Psychometrika

CONTENTS

VOLUME SEVENTEEN DECEMBER 1952 NUMBER 4

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

METHODOLOGY IN PSYCHOLOGY*

JOHN C. FLANAGAN

AMERICAN INSTITUTE FOR RESEARCH
UNIVERSITY OF PITTSBURGH

Introduction

An occasion of this sort seems a good time to take stock of the aims and methods of the members of this Society in their professional activities. Too often, even in a selected group of this type, a casual observer might get the impression that the aims of the members were primarily to prepare and get published a respectable number of journal articles with an occasional chapter in a book, and now and then a major publication of some type. Without wishing to lay claim to having made a research study in motivation, casual observation suggests that at least some of this activity is undertaken in the hope that individuals will be moved up to the next rung on the ladder or offered a more attractive position in some other locality.

In spite of such appearances it seems certain that anyone who has listened closely to the meetings both formal and informal during the past week would be forced to the conclusion that the individual members of this Society have a genuine interest in achieving more basic goals than those just mentioned.

As a basic factor to be considered in such a review as I propose, it is of importance that the overwhelming majority of the members of the Psychometric Society are individuals who regard themselves as primarily psychologists. It is true that they are specialists and a number of them have had special training in mathematics. But considered one by one as we run down the membership list, the impression seems unmistakable that their goals are psychological goals to a greater extent than mathematical goals. The interests of this group may, therefore, be regarded as both rational investigation in the mathematical field, especially in regard to statistical theory, and also empirical studies utilizing the methods of science. Both inductive and deductive methods play important roles in the activities of most of the members of the group. In the present discussion I wish to confine myself primarily to the scientific and psychological aspects of the activities of the Society members rather than the mathematical.

*This paper was presented as the Presidential Address to the Psychometric Society, September 5, 1952.

As psychologists, what are we trying to do and how are we going about doing it? Perhaps we should begin with some basic points on which we can hope to get general agreement. As our first statements let us take: The aim of psychology is the formulation of scientific knowledge regarding human behavior. All knowledge is based on perceptions made by individuals. It is assumed that observations give us a first approximation to knowledge. A primary function of science is to supply devices for increasing the precision of these initial observations. Following the lead of the logical positivists in philosophy and the operationists in physics, a number of psychologists, including S. S. Stevens, Kenneth Spence, Melvin Marx, and others, have worked on formalizing the bases for a modern methodology for psychology.

It is generally agreed that the ultimate aim of psychology is the understanding and explanation of behavior. The practical side of such understanding and explanation is the ability to predict and to control. Our ability to predict or influence human behavior depends on the development of principles or laws. Current theorists regard no such principles or laws as certain but only probable. Although empirical observation may enable us to make predictions which are later confirmed, the generalizations on which these predictions are based must be supplemented by more general theoretical principles or explanations which have been tested and confirmed before these generalizations can be admitted to the body of scientific knowledge.

These principles or explanations are derived from a propensity to make inferences of the kind consonant with such principles. The principles are usually made explicit only by reflecting on these inferences. The usual argument for the acceptance of psychological principles or laws is that they are the simplest explanations consistent with the observations which are available.

To go back a step, we find that generalizations are based on concepts or classes. The operationists regard a concept as defined by a set of operations. It is clear that concept formulation is essential to scientific studies, since all objects and events are unique to some extent. In setting up a class of objects or events, a number of criteria are usually established to determine membership in the class. Empirical classes of this type are never logically precise. Borderline cases can always be found which will defy any reasonable set of criteria. The concept can be simple and quantitative such as response time to a specified type of signal, or it can be complex and qualitative such as fear reaction to a particular threat or danger.

The psychologist must make judgments of sameness or of discrimination in studying a concept or a class. Another type of judgment must be made by the psychologist in setting up a criterion for accepting or rejecting objects or events as members of the class. This is a judgment of the relevance or importance of various associated characteristics and conditions. For example, brightness of the signal might be quite important in studying response time,

but relative humidity might be of negligible relevance. Often, of course, these judgments consist of hypotheses which, in other experiments, become experimental variables.

To summarize, then, the basic ingredients of scientific study are observation, concept formulation, generalization, and explanation. Let us turn then to the problem of the most effective methodology to get us from our original observations to explanations in terms of principles and laws.

Methodological Considerations for Psychological Problems

It seems useful at this point to formalize some of the basic considerations that are regarded as essential to sound methodology in psychology. For purposes of the present discussion, these will be discussed under seven main headings as follows:

1. *Defining and formulating problems*

The principal considerations in this regard were mentioned above and relate to the role of judgment in the development of concepts and classes. It should be clear from the discussion above that it is quite impossible to develop scientific knowledge without making judgments. It is desirable to utilize judgments which will provide as uniform conclusions from one researcher to another as is possible. It is believed that judgments of sameness fulfill this condition to a substantial degree and should be preferred.

Judgments of sameness may be used in setting up the concepts, classes, and series which are essential to collecting data for use in analysis. It is clearly impossible to make a statistical study of a group of psychotics. On the other hand, it is possible to make such studies of the weight, the age, and the ability to answer a series of questions of the members of this group. Some attribute of the members of a group or a sample must be clearly specified for study. It gives us no scientific knowledge to merely study people without defining what we are studying about them.

Another matter of importance in defining and formulating problems is that all verbal definitions are to some extent ambiguous because of the lack of precision in language. To achieve specificity in defining the aptitude or class to be studied and the function involved, it is desirable that these be stated in terms of the operations to be performed in selecting the observations which belong in the series or classes and in measuring or classifying each with respect to the attribute involved. These definitions should be such that all reasonable individuals perform essentially the same operations in the same way, and therefore obtain very similar results.

2. *Conditions and control*

This consideration owes its importance to the fact that it is impossible to duplicate any situation precisely. To obtain scientific knowledge it is

necessary that those conditions which are relevant to a significant degree to the outcome of the results be controlled in such a way that they will not cause the researcher to make incorrect inferences. In dealing with people, such conditions of motivation, attention, and set are usually very relevant and frequently particularly difficult to control.

3. *Observing and perceiving*

The fundamental consideration here is that of directed observations of the precise attributes to be observed and classifying the facts perceived with maximum accuracy. No two observers ever see exactly the same thing. Furthermore, their previous experiences differ in ways which affect their perceptions of phenomena observed. It is therefore essential that the precise operations to be performed in observing and in classifying what is observed be clearly specified if the observations are to be public and objective and are to be such that the results can be verified by other investigators.

4. *Recording and communicating*

This is an important consideration because memories tend to be dim, vague, and sometimes distorted images of perception. Language, which is our sole basis for communicating observations, can never provide a precise report. Verbal reports of perception tend to be incomplete, biased, and incoherent. This indicates the importance of immediate recording and also of immediate judgments as to the sameness of the objects, events, and attributes involved and the relevance of other factors in the situation in influencing the attributes being observed.

5. *Sampling*

This is a consideration very familiar to this group. The usual aim is to obtain the type of sampling conditions which underlie the theoretical sampling variations provided by standard formulas. The basic consideration is to insure that every member of the class being studied has equal opportunity to be included in the sample. Unfortunately, practical working conditions frequently prevent this from being the case. A great deal of the problem, therefore, is concerned with estimating the effects of these deviations from theoretical sampling conditions. One procedure which has been found very effective in handling sampling problems is the repeating of a study in its entirety to obtain an empirical estimate of sampling fluctuation. Another effective procedure is to bracket the results by obtaining an upper and lower limit based on samples which represent opposite extremes of the biases or determining factors present.

6. *Analyzing data*

The principal considerations in this area are those of accurate description and efficient comparison. It will be obvious from the preceding remarks that

quantitative and logical categories describing observations are of great importance in facilitating analysis.

7. *Interpreting results*

The chief factors in this area are the acceptance and rejection of hypotheses and similar problems regarding the degree of confidence to be given to the specific inferences or generalizations. The problems are primarily logical. It is important in using theoretical values to make allowances if theoretical conditions are not fulfilled. It is also desirable to avoid the error of accepting an alternate hypothesis merely because the data rejected the hypothesis being tested. Another important consideration is that the nature of the experimental studies and probabilities be kept in line and probabilities not be accepted as certainties but merely as working theories.

It is hoped that this group is in general agreement with the basic methodological considerations outlined above. Stated in general terms, they sound familiar and acceptable. Using these as a basis, what conclusions do we reach regarding our practical research problems? At this point I would like to narrow the discussion to one specific type of problem which I believe is most deserving of attention from psychologists at the present time. Specifically, these are problems with very important social implications. This does not mean that I prefer applied research to basic research. There seems adequate evidence at the present time that work on applied problems can lead to the discovery of basic scientific knowledge. The terms in which the problem is formulated and the general research methods used are the determining factors in whether a research study yields a specific result or one which may be generalized.

In the remarks that follow I am *not* proposing to tell specialists in learning, clinical, and social psychology how to carry out their research studies. These specialists have been trained in particular methods and habits of thinking and it does not seem appropriate to urge them to abandon these procedures in favor of what to me seems a more promising approach. I *am* recommending that members of this Society give serious consideration to certain methods for studying problems including those in the fields of learning, clinical, and social psychology.

The members of this Society have special training and experience in dealing with statistics from large samples. The methods of statistical analysis using computing and tabulating machines make it possible for persons with appropriate training to detect small differences even though they are obscured by the presence of a large number of disturbing factors. Similarly, the members of this Society have worked extensively with problems of testing involving the collection of a large number of observations on each individual in the sample. Not only is the number of observations usually large, but these observations frequently include a wide variety of attributes of the individual's performance. Because of the relevance of this type of experience

for effective work on the important social problems mentioned above, it is suggested that we leave the work with small sample theory to those psychologists making laboratory studies of rats, psychotics, and infants.

My specific recommendation, then, is that in their research work the members of this Society first select a broad, practical area containing important unsolved psychological problems and then use the approach and procedures described below to carry out research studies which will not only have immediate practical utility, but also contribute to our store of useful scientific knowledge.

Guiding Principles for Specific Steps

The specific approach and procedures recommended are discused under the following five headings: (1) Defining and formulating the specific problem, (2) Designing the study, (3) Collecting the data, (4) Analyzing the data, and (5) Interpreting and reporting the results.

1. *Defining and formulating the specific problems*

Having selected the broad, practical area, the following guiding principles are proposed for defining and formulating the specific problem. First, it is suggested that the problem be related to human activities which are practical, general, and normal, and not artificial, special, and abnormal. This suggests concentration on the everyday problems of human beings. It further suggests a selection of the more general problems encountered by normal people rather than working on the unusual or special case. Another important point to be considered in defining the problem is to establish the general aim of the people involved in this type of situation or event. Without a knowledge of the individual's goals, what he is trying to get to or get from, or what the individual is trying to do, his intentions, it is extremely difficult to formulate useful descriptions of his behavior. It should be noted that this type of problem will tend to be stated in broad and general rather than narrow and specific terms at the outset. The problem can be expected to be stated more as a program for research than as a specific study. Many specific problems for investigation will be developed as the study progresses.

2. *Designing the study*

The principles proposed in connection with preparing the design for the study are centered particularly around two concepts. The first of these is a preference for studying the problem, at least initially, in its natural or real setting rather than in an artificial or laboratory situation. This procedure has obvious advantages in the relevance and the validity of the results obtained. It also makes it especially easy to make use of persons in collecting the data who have had substantial experience in the situation being studied. Some of the difficulties of such a procedure are the lack of standardization in

various examples of the situation being investigated. Only recently has substantial progress been made on the problem of standardizing or adjusting for differences in conditions, and on producing uniformity in the perception of the goals or the general sets of the individuals with respect to the situation.

The second principle proposed in designing the study is the systematic collection of a large representative sample of events or behaviors of the type being studied. The laboratory tradition in psychology has led to the design of studies including only a small number of observations because of the difficulties of processing large numbers of cases in the laboratory. If useful results are to be obtained from small samples, great care must be used in reducing experimental error by controlling as many as possible of the factors which might obscure the effects of the variable being studied. It seems especially important that a very large number of observations in the natural setting be obtained to provide the basis for the tentative generalizations and hypotheses which are needed as a basis for establishing causal principles. Too often, specific studies have been set up to investigate problems formulated solely on the basis of limited self-observation or hunches.

3. *Collecting the data*

There are two important phases to collecting data. The first of these involves the process of observing and classifying the relevant aspects of the situation. It is proposed that only very simple judgments be required of the observers. Insofar as the observer can direct his attention to such operations as counting, reading along a scale, making judgments of "same" or "different," or judgments of "greater than" or "less than," it can be expected that little distortion will be introduced into the results by the observational process itself. In many practical situations it is essential that simple inferences be made by the observer. Under these circumstances, it is important that the observers be capable of making the specific types of inferences and judgments required. It appears especially undesirable to collect opinions, inferences about causal reasons, and complex judgments regarding the appropriateness and the quality of behavior in situations where simpler types of data are available. It is much easier to obtain agreement of independent observers as to whether or not a specific act was performed than as to whether this particular act indicated good adjustment to the situation. The over-all clinical type of estimate or judgment is the least satisfactory type of basic data for research studies of the sort proposed here.

The second phase of data collection is recording and reporting. It is futile to make precise observations if these are to be seriously distorted in the form in which they are reported. Immediate, on-the-spot recording by the observer is the most satisfactory procedure for collecting data. In situations where memory images must be used it is desirable that steps be taken to check the observations at the time they are made, and to reinforce them by

recall and rehearsal shortly afterwards. The observer is also important in this phase. Insofar as perceptions and inferences may be distorted by previous experience in the specific type of situation, the observer should be selected and trained to minimize the introduction of bias from this source.

4. *Analyzing the data*

The fundamental principle governing the analysis of data is that the sole purpose of this step is the more efficient description of these data. No really new information is added during this process. All of the information must be inherent in the data as collected. Summaries and descriptive statistics frequently make it easier to see relationships and differences in the data, and they of course provide the basis for tests of significance. However, as abstractions, they always contain less information than the originally collected data. Once the data have been collected, nothing can be done with them to improve the comprehensiveness, specificity of detail, or validity of the information they provide.

In many laboratory studies the data are so consistent and the experimental error so small that little in the way of statistical analysis is required to reveal the essential relationships. In the type of study proposed here this is not the case. Unusual effort and skill are required to discover and describe the fundamental relationships which are concealed within the large mass of the initial data. All of the tricks of multivariate analysis including factor analysis, partial correlation, and discriminant analysis will have to be used for effective handling of these problems.

5. *Interpreting and reporting the results*

Although most of the problems involved in interpreting and reporting results are common to all types of research, there are certain aspects which seem to deserve special emphasis here. The first responsibility of the investigator is to describe precisely the problem selected for investigation. Similarly, if the observations were not made on a random sample of human beings, the nature of the group should be reported as specifically as possible. Since it is usually hoped that the results can be generalized to other groups, any limitations imposed by the nature of the specific group used should be brought into clear focus. In similar fashion, the procedures in each of the other steps listed above should be clearly described so that the decisions of the investigator in carrying out the various steps of the study can be reviewed and evaluated by those interested in the results. It is important that limitations of a study be clearly reported. It is also essential that the research worker make available to others his considered judgment regarding the degree of credibility which should be attached to these findings. This is frequently a difficult type of judgment to make, but it is usually one the investigator is better prepared to make than his colleagues who are interested in utilizing his findings.

Examples of Applications

By way of conclusion, some of the implications of the above remarks for rather specific problems will be mentioned. Probably the best known application of the approach and the procedures described above is in connection with job analysis. In this area much progress has been made in substituting systematically collected factual data for the opinions, hunches, and general impressions of various types of observers. The critical incident technique and related procedures for collecting reports of observations made in accordance with detailed instructions and criteria have produced uniformly excellent results in the hands of personnel trained in their use. Detailed statements of the requirements for industrial plant employees, office workers, dentists, infantrymen, combat leaders, aircrew members, research workers, and many other groups are now available.

The detailed statements of job requirements obtained in the type of study described in the preceding paragraph have led directly to the development of new procedures for constructing selection tests, proficiency measures, and criteria of job performance. The importance of objective procedures, simple judgments, frequent recording, and precise criteria, have been demonstrated for a wide variety of types of measures in these areas.

Not so well known are recent studies of the learning process utilizing this general approach. Following the procedures described above, studies of learning have centered around two of the most common learning situations. The first of these is the teacher-student situation. The activities of the teacher in the typical school situation which promote learning or interfere with it have been studied by a number of investigators following the general methods outlined here. A similar study has been made of pilot instructors. In this situation all of the flight instruction is given individually. Since motivation is usually high and the opportunities for acquiring skill other than in the instructor-student situation are small, this provides an excellent opportunity to study the instructional process. The preliminary results of these studies suggest that this approach has substantial promise as an aid to gaining a better understanding of the problems of learning.

The second type of study on the learning problem using this general approach consists of a systematic analysis of the process of getting information from written materials. Although this study is in an even more preliminary stage, it seems likely to make a definite contribution to the practical problems of human learning.

A few studies have also been made now applying this methodology to problems of clinical psychology. Perhaps the best example to date is the development of a tentative definition in behavioral terms of immaturity reaction. This study was recently completed by Dr. Leo Eilbert. The acceptability of his detailed behavioral definition to a panel of fifteen psychiatrists to whom it was submitted for review is interpreted as an indication

of great promise for the practical usefulness of the results which can be expected from applying these procedures on a larger scale to other groups showing unusual behavior. Work has also been done on studies of the therapeutic process. It is hoped that this general type of procedure will make it possible to replace much of the subjective opinion and impression in this field with data of a more objective and factual nature.

The last application proposed is to the problems of social psychology. This field also seems to be replete with studies of opinions, ratings, impressions, and theories stated in such general terms that they cannot be tested. It is believed that many of the studies of reported attitudes, expressed preferences, estimated motivations, and role judgments can be replaced by reports of specific observed behavior with great profit to the advancement of knowledge in this field.

Before closing, let me reemphasize the fact that most if not all of the principles discussed here are both well known and regularly used by many of the members of this group. This presentation has attempted to formalize and underline some of the more important aspects of this approach and to try to encourage you and your students to exploit more fully the training and experience which provide a sound basis for making substantial progress in solving urgent practical problems and at the same time increasing our store of scientific knowledge regarding human behavior.

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

A STATISTICAL DESCRIPTION OF VERBAL LEARNING*

GEORGE A. MILLER AND WILLIAM J. MCGILL

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Free-recall verbal learning is analyzed in terms of a probability model. The general theory assumes that the probability of recalling a word on any trial is completely determined by the number of times the word has been recalled on previous trials. Three particular cases of this general theory are examined. In these three cases, specific restrictions are placed upon the relation between probability of recall and number of previous recalls. The application of these special cases to typical experimental data is illustrated. An interpretation of the model in terms of set theory is suggested but is not essential to the argument.

The verbal learning considered in this paper is the kind observed in the following experiment: A list of words is presented to the learner. At the end of the presentation he writes down all the words he can remember. This procedure is repeated through a series of n trials. At the present time we are not prepared to extend the statistical theory to a wider range of experimental procedures.

The General Model

We shall assume that the degree to which any word in the test material has been learned is completely specified by the number of times the word has been recalled on preceding trials. In other words, the probability that a word will be recalled on trial $n + 1$ is a function of k, the number of times it has been recalled previously. (Symbols and their meanings are listed in Appendix C at the end of the paper.)

Let the probability of recall after k previous recalls be symbolized by τ_k . Then the corresponding probability of failing to recall the word is $1 - \tau_k$. When a word has been recalled exactly k times on the preceding trials, we shall say that the word is in state A_k . Thus before the first trial all the words are in state A_0 ; that is to say, they have been recalled zero times on previous trials. Ideally, on the first trial a proportion τ_0 of these words is recalled and so passes from state A_0 to state A_1 . The proportion $1 - \tau_0$ is not recalled and so remains in state A_0 . On the second trial the

*This research was facilitated by the authors' membership in the Inter-University Summer seminar of the Social Science Research Council, entitled Mathematical Models for Behavior Theory, held at Tufts College, June 28–August 24, 1951. The authors are especially grateful to Dr. F. Mosteller for advice and criticism that proved helpful on many different occasions.

words that remained in A_0 undergo the same transformation as before. Of those in A_1, however, the proportion $1 - \tau_1$ is not recalled and so remains in A_1.

One general problem is to determine the proportion of words expected in state A_k on trial n. Let $p(A_k,n)$ represent the probability that a word is in state A_k on trial n. Since these are probabilities, they must sum to unity on any given trial:

$$\sum_k p(A_k, n) = 1.$$

The number of trials and the total number of times a word has been recalled must assume non-negative, integral values. We assume that a word can be recalled only once per trial at most, so the number of recalls cannot exceed the number of trials. Therefore, we have

$$p(A_k, n) = 0 \qquad \text{for} \qquad k < 0, n < 0, n < k.$$

We also assume that none of the words can have been recalled before the first trial, so for $n = 0$,

$$p(A_k, 0) = \begin{cases} 1 & \text{for} \quad k = 0, \\ 0 & \text{for} \quad k \neq 0. \end{cases}$$

For all trials we have the difference equation:

$$p(A_k, n + 1) = p(A_k, n)(1 - \tau_k) + p(A_{k-1}, n)\tau_{k-1}. \tag{1}$$

This equation reflects the fact that a word can get into state A_k on trial $n + 1$ in only two ways: (a) either it is in A_k on trial n and is not recalled on trial $n + 1$, or (b) it is in A_{k-1} on trial n and is recalled on trial $n + 1$.

The following rationalization for this scheme is in the spirit of the statistical theories of learning developed by Bush and Mosteller (1) and by Estes (3). The rationalization is not necessary for the development of the mathematics, but it gives an alternative way of thinking about the present model and helps to clarify its relation to the earlier theories. On the first presentation of the list of words a random sample of stimulus elements is conditioned to the appropriate response for each word. The measure of this set of conditioned elements is τ_0. (The total measure of the set of all stimulus elements for a given word is assumed to be unity, so the measure can be regarded as a probability.) If a word is not recalled, the measure of conditioned elements for that word is unchanged. But if a word is recalled, the proportion of conditioned elements is increased. The effect of recalling a word is to take another random sample of elements from the total set and to condition them. The proportion of elements conditioned when a word in state A_k is recalled is $\tau_{k+1} - \tau_k$. More precise interpretation of this set-theoretical argument will be presented when we consider the special cases of the general theory.

The general solution of (1) when all the τ_k are different is (see Appendix A):

$$p(A_0, n) = (1 - \tau_0)^n, \qquad \text{for} \qquad k = 0,$$

$$p(A_k, n) = \tau_0\tau_1 \cdots \tau_{k-1} \sum_{i=0}^{k} \frac{(1 - \tau_i)^n}{\prod_{\substack{j=0 \\ j \neq i}}^{k} (\tau_j - \tau_i)}, \qquad \text{for } k > 0. \tag{2}$$

The denominator of each of the fractions in the summation includes all differences of the form $(\tau_j - \tau_i)$ except for the zero difference $(\tau_i - \tau_i)$.

The expected number of times a word is recalled, all told, up to and including trial n, is, by definition,

$$E(k, n) = \sum_{k=0}^{n} kp(A_k, n). \tag{3}$$

The expected proportion of words recalled on trial $n + 1$ is the difference, $E(k, n + 1) - E(k, n)$, between the cumulative values on successive trials. This difference is the theoretical recall score and we symbolize it by ρ_{n+1}. Thus we have the general relation

$$\rho_0 = 0, \qquad \text{for } n = 0,$$

$$\rho_{n+1} = E(k, n + 1) - E(k, n), \qquad \text{for } n + 1 > 0. \tag{4}$$

An alternative expression for ρ_{n+1} can be obtained as follows. On trial n the probability that a word is in state A_k is $p(A_k, n)$. The probability of recall in state A_k is τ_k. The product $\tau_k \cdot p(A_k, n)$ is, therefore, the probability that a word will both be in A_k on trial n and also be recalled on trial $n + 1$. If these joint probabilities are summed over all the states A_k from $k = 0$ to $k = n$, we have the total probability that a word will be recalled on trial $n + 1$. That is to say, we have ρ_{n+1}:

$$\rho_{n+1} = \sum_{k=0}^{n} \tau_k p(A_k, n). \tag{5}$$

The two expressions (4) and (5) are equivalent, which can be shown as follows. From (3) and (4) together we have

$$\rho_{n+1} = \sum_{k=0}^{n+1} kp(A_k, n + 1) - \sum_{k=0}^{n} kp(A_k, n).$$

The first summation on the right can be rewritten by substituting for $p(A_k, n + 1)$ according to (1):

$$\sum_{k=0}^{n+1} kp(A_k, n + 1) = \sum_{k=0}^{n+1} kp(A_k, n)(1 - \tau_k) + \sum_{k=0}^{n+1} kp(A_{k-1}, n)\tau_{k-1}$$

$$= \sum_{k=0}^{n} kp(A_k, n) - \sum_{k=0}^{n} kp(A_k, n)\tau_k + \sum_{k=0}^{n} (k + 1)p(A_k, n)\tau_k.$$

When this result is substituted into the expression for ρ_{n+1}, we have

$$\rho_{n+1} = -\sum_{k=0}^{n} kp(A_k, n)\tau_k + \sum_{k=0}^{n} (k+1)p(A_k, n)\tau_k$$

$$= \sum_{k=0}^{n} \tau_k p(A_k, n),$$

which is the desired result.

The asymptotic behavior of the model as n increases without limit can be deduced from the general solution (2). First consider the case in which one or more of the transitional probabilities τ_k is zero. All the words start in state A_0 and have a positive probability of moving along to states A_1, A_2, etc., up to the first state, A_h, with zero transitional probability, $\tau_h = 0$. There the words are trapped; eventually all the words are recalled exactly h times and cannot be recalled again. This fact can be seen from (2): If $\tau_i > 0$, then all the terms $(1 - \tau_i)^n$ in (2) go to zero as $n \to \infty$. Thus $p(A_k, n)$ goes to zero for $k < h$. For $k > h$, the product in front of the summation must include $\tau_h = 0$, and so $p(A_k, n) = 0$ for $k > h$. When $k = h$, however, $(1 - \tau_h)^n = (1 - 0)^n = 1$, and so this term in the summation of (2) does not go to zero. Instead, when $\tau_h = 0$ and $\tau_i > 0$ for $i < h$,

$$\lim_{n\to\infty} p(A_h, n) = \frac{\tau_0\tau_1 \cdots \tau_{h-1}}{(\tau_0 - \tau_h)(\tau_1 - \tau_h) \cdots (\tau_{h-1} - \tau_h)} = 1.$$

The recall score, ρ_{n+1}, then approaches zero as an asymptote; from (5),

$$\lim_{n\to\infty} \rho_{n+1} = \sum_{k=0}^{\infty} \tau_k [\lim_{n\to\infty} p(A_k, n)] = 0,$$

since the probability at the asymptote is concentrated at state A_h, and for this state $\tau_h = 0$. This case is of little interest for an acquisition theory, since the asymptote of the learning curve is at zero. Therefore, in what follows, we shall be concerned only with the case in which all the τ_k are different and greater than zero.

If all the transitional probabilities τ_k are greater than zero, then from (2) we see that as n approaches infinity all the terms in the summation go toward zero for all finite values of k. Consequently the sum of the $p(A_k, n)$ can be made as near zero as we please for any finite k by selecting a large enough value of n. In the limit, therefore, the probability of any finite number of recalls is zero. Since the sum of the $p(A_k, n)$ must equal unity, almost all the probability comes to be concentrated in state A_∞ and we have for the limit when all $\tau_k > 0$,

$$p(A_\infty, \infty) = 1.$$

We are now able to show that a word in state A_k has probability one of moving to state A_{k+1}, if the learning process is continued indefinitely. This happens because almost all words eventually reach state A_∞. Thus we can write, for the probability of leaving state A_k on some trial,

$$\sum_{n=k}^{\infty} \tau_k p(A_k, n) = 1,$$

or,

$$\sum_{n=k}^{\infty} p(A_k, n) = \frac{1}{\tau_k} \qquad \text{for } \tau_k > 0.$$

In all the cases we shall consider in this paper the value of τ_k will approach an asymptote as $k \to \infty$. We are interested in placing the following restrictions on the τ_k:

$$\tau_k \neq \tau_j,$$

$$\tau_k > 0,$$

$$\lim_{k\to\infty} \tau_k = m \leq 1.$$

The first two conditions insure that $p(A_k, n)$ goes toward zero for finite k and large n. The third condition provides the asymptotic value of τ_k for infinite k. In the summation for the limiting value of ρ_{n+1}, all terms are zero out to infinity, and so we have

$$\lim_{n\to\infty} \rho_{n+1} = mp(A_\infty, \infty) = m. \tag{5'}$$

In other words, if we assume that m is the asymptotic value of τ_k as $k \to \infty$, then m is also the asymptotic value of ρ_{n+1} as $n \to \infty$.

In the special cases discussed below, a restriction is placed upon the value of τ_k in the form of the linear difference equation,*

$$\tau_{k+1} = a + \alpha\tau_k, \tag{6}$$

where $0 \leq a \leq 1$ and $0 \leq \alpha \leq 1 - a$. The limits for α have been chosen so that τ_{k+1} is bounded between zero and one and, since we are interested in acquisition, so that $\tau_{k+1} \geq \tau_k$.

Consider the following development of (5):

$$\rho_{n+2} = \sum_{k=0}^{n+1} \tau_k p(A_k, n+1).$$

*We have tried to observe the convention that parameters are represented by Greek letters and statistical estimates are represented by Roman letters. In the case of a and m, however, we have violated this convention in order to make our symbols coincide with those used by other workers. The symbols m, a, α, and p were originally proposed by Bush and Mosteller.

Now substitute for $p(A_k , n + 1)$ according to (1):

$$\rho_{n+2} = \sum_{k=0}^{n+1} \tau_k p(A_k , n)(1 - \tau_k) + \sum_{k=0}^{n+1} \tau_k p(A_{k-1} , n)\tau_{k-1}$$

$$= \rho_{n+1} - \sum_{k=0}^{n} \tau_k^2 p(A_k , n) + \sum_{k=0}^{n} \tau_{k+1}\tau_k p(A_k , n).$$

Next we substitute for τ_{k+1} according to (6):

$$\rho_{n+2} = \rho_{n+1} - \sum_{k=0}^{n} \tau_k^2 p(A_k , n) + \sum_{k=0}^{n} (a + \alpha\tau_k)\tau_k p(A_k , n)$$

$$= (1 + a)\rho_{n+1} - (1 - \alpha) \sum_{k=0}^{n} \tau_k^2 p(A_k , n)$$

$$= (1 + a)\rho_{n+1} - (1 - \alpha)E(\tau_k^2 , n + 1), \tag{7}$$

where $E(\tau_k^2 , n + 1)$ is the second raw moment of the τ_k (as ρ_{n+1} is the first raw moment) for trial $n + 1$.

Restriction (6) brings the system into direct correspondence with a special case of the theory developed by Bush and Mosteller. In their terminology, an operator Q_1 is applied to the probability of response, p, to give $a_1 + \alpha_1 p$ as the new probability whenever a trial is successful. A second operator Q_2 is applied to give $a_2 + \alpha_2 p$ whenever a trial is unsuccessful. In the present application of this more general theory, Q_1 is preserved intact by restriction (6), but Q_2 is assumed to be the identity operator. That is to say, a_2 is zero and α_2 is unity, so $Q_2 p = p$. In the present application, an unsuccessful trial consists of the omission of the word during recall. It seems reasonable to assume that the non-occurrence of a word has no effect upon its probability of occurrence on the next trial. How successful this simple assumption is will be seen when we examine the data.

Analysis of the Data

At the end of the experiment the experimenter has collected a set of word lists—the words recalled by the learner on successive trials. These recall lists will usually contain a small number of words that did not occur in the presentation. These spontaneous additions by the learner are of some interest in themselves, but we shall ignore them in the present discussion.

We would like to use the data contained in the word lists to obtain an estimate of ρ_{n+1} in (5). We shall refer to the estimate as r_{n+1}. There are, we suppose, N words provided by the experimenter as learning material in the experiment. It seems reasonable to assume that under certain conditions these words are homogeneous. By this we imply that the responses to all of the words in state A_k may be considered as estimates of the same transitional probability of recall, τ_k.

We can then define a convenient statistic,

$$r_{n+1} = \frac{1}{N} \sum_{k=0}^{n} \sum_{i=1}^{N} X_{i,k,n+1} \cdot \tag{8}$$

The numbers, $X_{i,k,n+1}$, are either zero or one. The subscripts k and $n + 1$ have the same meaning that we have attached to them previously. They indicate that we are looking at an event that occurs on trial $n + 1$ to a word in state A_k. The first summation is carried out over i, the experimental words, with k fixed to show that we count the number of words in each state. The rules that determine whether an $X_{i,k,n+1}$ is zero or one are straightforward. The $X_{i,k,n+1}$ are zero for all words not in state A_k when summing on i. They are zero for any word in state A_k, if a recall fails to occur on trial $n + 1$. Lastly the $X_{i,k,n+1}$ are 1 for any word in state A_k, provided that a recall occurs on trial $n + 1$. The second summation extends over k, the various states. This summation goes only up to n because our reference point for determining the number of states is trial n. These rules determine r_{n+1} as the proportion of correct responses to the N experimental words on trial $n + 1$.

To show that r_{n+1} is unbiased we observe that

$$E(r_{n+1}) = \frac{1}{N} \sum_{k=0}^{n} \left[E\left(\sum_{i=1}^{N} X_{i,k,n+1} \right) \right].$$

The expectation of any $X_{i,k,n+1}$ in state A_k is τ_k. Thus the expectation of the sum in the brackets is $N \cdot \tau_k \cdot p(A_k , n)$. Substituting this into the expression for $E(r_{n+1})$, we find

$$E(r_{n+1}) = \sum_{k=0}^{n} \tau_k p(A_k , n),$$

$$E(r_{n+1}) = \rho_{n+1} \cdot \tag{9}$$

The sampling variance of r_{n+1} around ρ_{n+1} is determined by the variances of the various $X_{i,k,n+1}$ around the transitional probabilities, τ_k.

$$\text{Var}\,(r_{n+1}) = \frac{1}{N^2} \sum_{k=0}^{n} \text{Var}\left(\sum_{i=1}^{N} X_{i,k,n+1} \right).$$

The variance of any $X_{i,k,n+1}$ in state A_k is binomial and is given by $\tau_k(1 - \tau_k)$. The variance of $\sum_{i=1}^{N} X_{i,k,n+1}$ thus becomes $N\, p(A_k , n)\tau_k\,(1 - \tau_k)$. Substituting this into the expression for Var (r_{n+1}), we obtain

$$\text{Var}\,(r_{n+1}) = \frac{1}{N} \sum_{k=0}^{n} p(A_k , n)\tau_k(1 - \tau_k). \tag{10}$$

It should be noted that this variance is never larger than the binomial variance

$$\frac{1}{N}\, \rho_{n+1} \cdot (1 - \rho_{n+1}),$$

since the binomial variance includes in addition to (10) a term that depends on the variance of the τ_k around ρ_{n+1} ,

$$\text{Var}\,(r_{n+1}) = \frac{\rho_{n+1}(1 - \rho_{n+1})}{N} - \frac{1}{N}\left\{ \sum_{k=0}^{n} \tau_k^2 p(A_k\,,\, n) - \rho_{n+1}^2 \right\}. \qquad (10')$$

In order to apply the general theory we must obtain estimates of the transitional probabilities, τ_k . Now τ_k is the probability of moving from state A_k to A_{k+1} and is assumed to be constant from trial to trial. After trial n a certain number of words, $N_{k,n}$, are in state A_k . Of these $N_{k,n}$ words, some go on to A_{k+1} and some remain in A_k on trial $n + 1$. The fraction that moves up to A_{k+1} provides an estimate of τ_k on that trial. Therefore, on every trial we obtain an estimate of τ_k . Call these estimates $t_{k,n+1}$. Then

$$t_{k,n+1} = \frac{\sum_{i=1}^{N} X_{i,k,n+1}}{N_{k,n}} .$$

If $N_{k,n}$ is zero, no estimate is possible.

Next we wish to combine the $t_{k,n+1}$ to obtain a single estimate, t_k , of the transitional probability, τ_k . The least-squares solution, obtained by minimizing $(t_{k,n+1} - \tau_k)^2$, is the direct average of the $t_{k,n+1}$. This estimate is unbiased, but it has too large a variance because it places undue emphasis upon the $t_{k,n+1}$ that are based on small values of $N_{k,n}$. We prefer, therefore, to use the maximum-likelihood estimate,

$$t_k = \frac{\sum_n N_{k,n} t_{k,n+1}}{\sum_n N_{k,n}}\,, \qquad (11)$$

which respects the accuracy of the various $t_{k,n+1}$.

For example, after trial 7 there may be 10 words in state A_3 . Of these 10, 6 are recalled on trial 8. This gives the estimate $t_{3,8} = 6/10$. Every trial on which $N_{3,n} \neq 0$ provides a similar estimate, $t_{3,n+1}$. The final estimate of τ_3 is obtained by weighting each of these separate estimates according to the size of the sample on which it is based and then averaging. This procedure is repeated for all the τ_k individually as far as the data permit.

The $t_{k,n+1}$ are also useful to check the basic assumption that τ_k is independent of n. If the $t_{k,n+1}$ show a significant trend, this basic assumption is violated.

The Simplest Case: One Parameter

The computation of $p(A_k\,,\, n)$ from (2) for the general case is exceedingly tedious as n and k become moderately large. We look, therefore, for a simple relation among the τ_k of the form of restriction (6). The first case that we

shall consider is

$$\tau_0 = a,$$
$$\tau_{k+1} = a + (1 - a)\tau_k . \tag{12}$$

In this form the model contains only the single parameter, a. The solution of the difference equation (12) is

$$\tau_k = 1 - (1 - a)^{k+1}. \tag{13}$$

The interpretation of (13) in set-theoretical terms runs as follows: On the first presentation of the list a random sample of elements is conditioned for each word. The measure of this sample is a, and it represents the probability, τ_0, of going from state A_0 to state A_1. If a word is not recalled, no change is produced in the proportion of conditioned elements. When a word is recalled, however, the effect is to condition another random sample of elements, drawn independently of the first sample, of measure a to that word. Since some of the elements sampled at recall will have been previously conditioned, after one recall we have (because of our assumption of independence between successive samples):

$$\begin{pmatrix}\text{Elements conditioned}\\ \text{during presentation}\end{pmatrix} + \begin{pmatrix}\text{Elements conditioned}\\ \text{during the recall}\end{pmatrix} - \begin{pmatrix}\text{Common}\\ \text{elements}\end{pmatrix}$$
$$= a + a - a^2 = 1 - (1 - a)^2.$$

This quantity gives us the transitional probability τ_1 of going from A_1 to A_2, from the first to the second recall. The second time a word is recalled another independent random sample of measure a is drawn and conditioned, so we have

$$\tau_2 = [1 - (1 - a)^2] + a - a[1 - (1 - a)^2] = 1 - (1 - a)^3.$$

Continuing in this way generates the relation (13).

With this substitution the general difference equation (1) becomes

$$p(A_k , n + 1) = p(A_k , n)(1 - a)^{k+1} + p(A_{k-1} , n)[1 - (1 - a)^k].$$

The solution of this difference equation can be obtained by the general method outlined in Appendix A or by the appropriate substitution for τ_k in (2). The solution is

$$p(A_0 , n) = (1 - a)^n,$$
$$p(A_k , n) = (1 - a)^{n-k} \prod_{i=0}^{k-1} [1 - (1 - a)^{n-i}]. \tag{14}$$

From definition (5) it is possible to obtain the following recursive ex-

pression for the recall on trial $n + 1$ (see Appendix B):

$$\rho_{n+1} = a + (1 - a)[1 - (1 - a)^n]\rho_n . \tag{15}$$

The variance of the recall score, r_{n+1} , is

$$\text{Var}\,(r_{n+1}) = \frac{1}{aN}\,(\rho_{n+2} - \rho_{n+1}). \tag{16}$$

In order to illustrate the application of these equations, we have taken the data from one subject in an experiment by J. S. Bruner and C. Zimmerman (unpublished). In their experiment a list of 64 monosyllabic English words was read aloud to the subject. At the end of each reading the subject wrote all of the words he could remember. The order of the words was scrambled before each reading. A total of 32 presentations of the list was given.

From the detailed analysis of the estimates of τ_k derived from this subject's data it was determined that a value of $a = 0.22$ would provide a

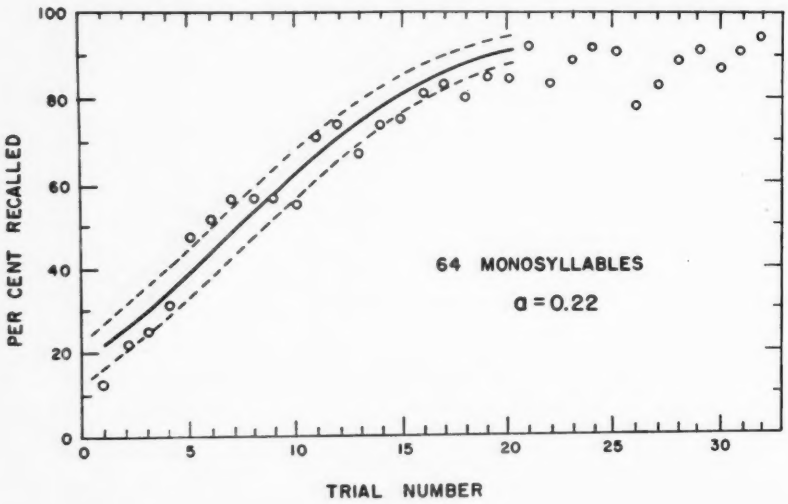


FIGURE 1

Comparison of Theoretical and Observed Values of ρ_n for the One-Parameter Case. Dotted line is drawn $\pm$ one standard deviation from ρ_n .

good fit. In Figure 1 the values of ρ_{n+1} computed from (15) are given by the solid function. The data are shown by the open circles. The dotted lines are drawn $\pm$ one standard deviation from ρ_{n+1} as computed from the variance in (16). The single parameter gives a reasonably adequate description of these data, at least through the first 20 trials. From the 20th trial on, however, it seems that the subject "forgets as fast as he learns." He seems to

reach an asymptote somewhat below the theoretical value at unity. The introduction of an asymptote less than unity will be discussed in connection with the three-parameter case.

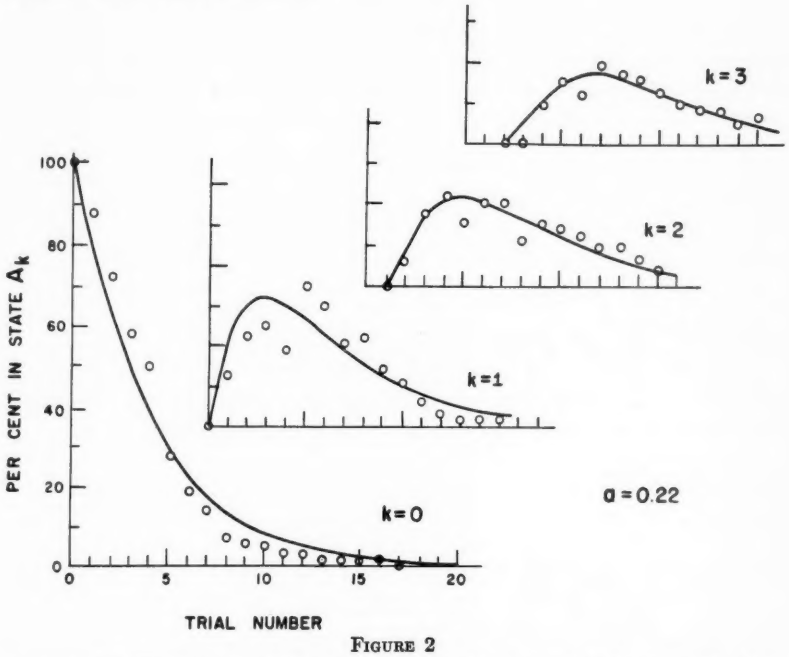


FIGURE 2

Comparison of Theoretical and Observed Values of $p(A_k, n)$ for the One-Parameter Case

As a further check on the correspondence of theory and data, Figure 2 shows the predicted and observed values of $p(A_k, n)$ as a function of n, for $k = 0, 1, 2, 3$.

Second Case: Two Parameters.

In the one-parameter form of the theory it is assumed that the proportion of elements sampled during the presentation of the list is the same as the proportion sampled during each recall. Most data are not adequately described by such a simple model. At the very least, then, it is necessary to consider the situation when these two sampling constants are different. In order to introduce the second parameter, we phrase restriction (6) in the following form:

$$\tau_0 = p_0 ,$$
$$\tau_{k+1} = a + (1 - a)\tau_k , \qquad (17)$$

where p_0 is the proportion of elements conditioned during the presentation. The solution of this difference equation can be written

$$\tau_k = 1 - (1 - p_0)(1 - a)^k. \tag{18}$$

On the first presentation of the list a random sample of measure p_0 is conditioned to every word. When a word is recalled, a random sample of measure a is drawn and conditioned. After one recall, therefore, the measure of conditioned elements is

$$\tau_1 = p_0 + a - ap_0 = 1 - (1 - p_0)(1 - a).$$

After two recalls the measure of conditioned elements is

$$\begin{aligned}\tau_2 &= [1 - (1 - p_0)(1 - a)] + a - a[1 - (1 - p_0)(1 - a)] \\ &= 1 - (1 - p_0)(1 - a)^2.\end{aligned}$$

Continuing in this way generates the relation (18).

With this substitution the general difference equation (1) becomes

$$\begin{aligned}p(A_k, n + 1) &= p(A_k, n)(1 - p_0)(1 - a)^k \\ &\quad + p(A_{k-1}, n)[1 - (1 - p_0)(1 - a)^{k-1}].\end{aligned} \tag{19}$$

The solution of (19) is

$$p(A_0, n) = (1 - p_0)^n,$$

$$p(A_k, n) = (1 - p_0)^{n-k} \prod_{i=0}^{k-1} \frac{[1 - (1 - p_0)(1 - a)^i][1 - (1 - a)^{n-i}]}{1 - (1 - a)^{i+1}}. \tag{20}$$

When $p_0 = a$, (20) reduces to (14).

The recursive form for the recall now becomes (see Appendix B)

$$\rho_{n+1} = p_0 + (1 - p_0)[1 - (1 - a)^n]\rho_n. \tag{21}$$

The variance of r_{n+1} is

$$\text{Var}(r_{n+1}) = \frac{1}{aN}(\rho_{n+2} - \rho_{n+1}). \tag{22}$$

In order to illustrate the application of these equations we have selected two sets of data. The first set was collected by Bruner and Zimmerman. A list of 32 monosyllabic words was read aloud. At the end of each reading the subject wrote all of the words he could remember. The order of the words was scrambled before every reading. A total of 32 presentations of the list was given.

From the analysis of the t_k calculated for this particular subject it was found that $a = 0.10$ and $p_0 = 0.27$ gave a good description of the data. In Figure 3 the values of ρ_{n+1} computed from (21) are shown by the solid function. The data are given by the open circles. The dotted lines are drawn

$\pm$ one standard deviation from ρ_{n+1} as computed from (22). As a further check, Figure 4 shows the predicted and observed values of $p(A_k\ ,\ n)$ as a function of n for $k = 0, 1, 2, 3$.

The distribution of cumulative recalls on any given trial provides still another way of viewing the data. In Figure 5, the cumulative distribution of k, the number of recalls, is shown for trials 5, 10, 15, 20. The proportion of test words recalled k times or less is plotted for comparison on each trial.

The second set of data was collected by M. Levine. He read aloud a 100-word anecdote. At the end of the reading, the subject wrote down all he could remember. Four such trials were given. The order of the words was not scrambled during the interval between trials.

From the analysis of the data for this particular subject it was found that $a = 0.87$ and $p_0 = 0.61$ gave a good description of the results. Figure 6 shows the comparison of theory and experiment both for ρ_{n+1} and for $p(A_k\ ,\ n)$ for $k = 0, 1, 2$.

As a general observation, we have noted that when the order of the words is not scrambled between trials, the parameter a is relatively large. This is to say, when the words are not scrambled, there is a much higher probability that the same words will be recalled on successive trials. This effect is related to the serial-position curve. The subject recalls words at the beginning and at the end of the list. If these words remain in their favored positions, they continue to be recalled. New words are added to those recalled at the ends at a rate determined by p_0, so the learning works from the two ends toward the middle, which is the last to be learned. This effect has been noted with lists of randomly selected English words as well as with anecdotes.

Third Case: Three Parameters

In the one- and two-parameter cases we have assumed that after sufficient practice the subject should eventually reach perfect performance. Some data, however, seem to evade this simple assumption and so it is necessary to consider what happens when a lower asymptote is introduced. Such a parameter may be necessary when, for example, the period of time allowed for recall is limited.

To introduce the third parameter we adopt the general restriction (6)

$$\tau_0 = p_0\ ,$$
$$\tau_{k+1} = a + \alpha\tau_k\ , \qquad \text{where} \qquad 0 \leq a \leq 1 - \alpha \leq 1. \tag{23}$$

The solution of (23) can be written

$$\tau_k = \frac{a}{1-\alpha} - \left(\frac{a}{1-\alpha} - p_0\right)\alpha^k. \tag{24}$$

When $\alpha = 1 - a$, (24) reduces to (18). From (24) we see that as k increases without limit, τ_k approaches $a/(1-\alpha)$ as an asymptote. From (5′) we know

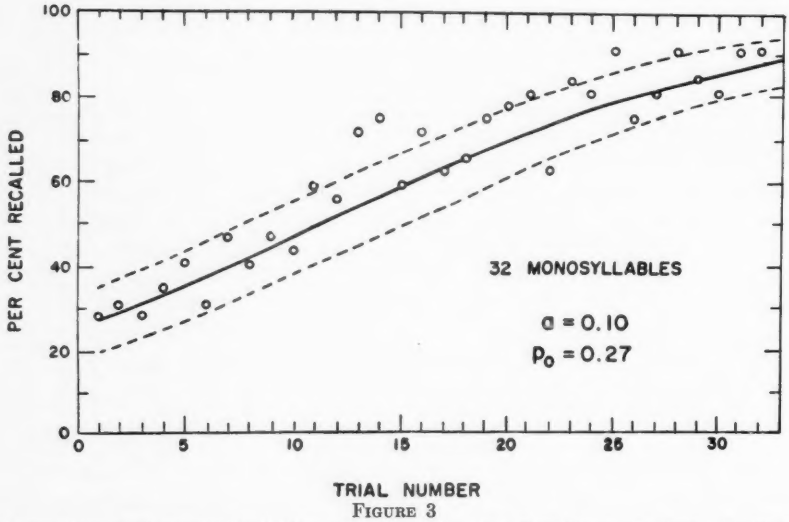


FIGURE 3

Comparison of Theoretical and Observed Values of ρ_n for a Two-Parameter Case. Dotted line is drawn $\pm$ one standard deviation from ρ_n.

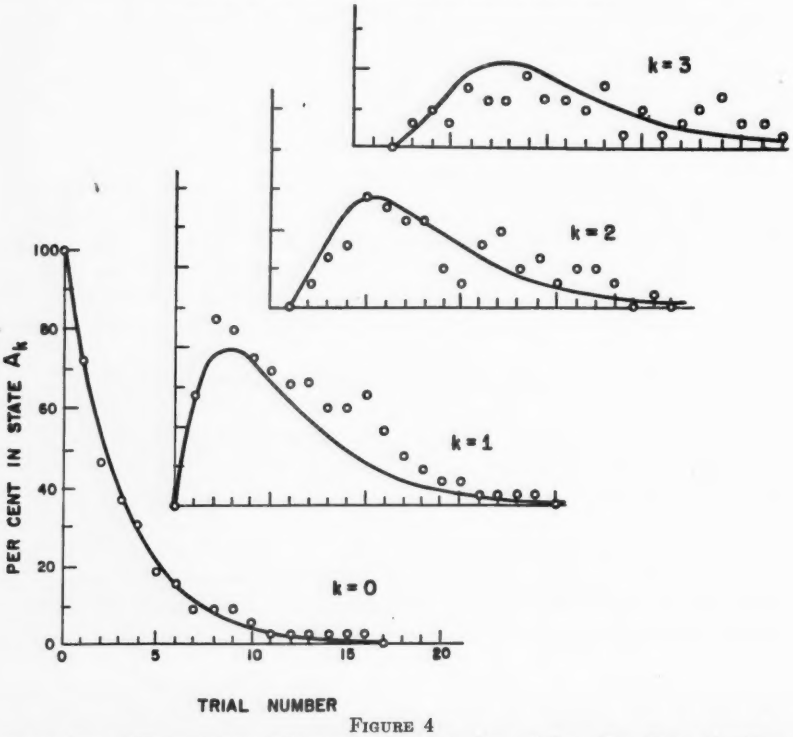


FIGURE 4

Comparison of Theoretical and Observed Values of $p(A_k, n)$ for a Two-Parameter Case

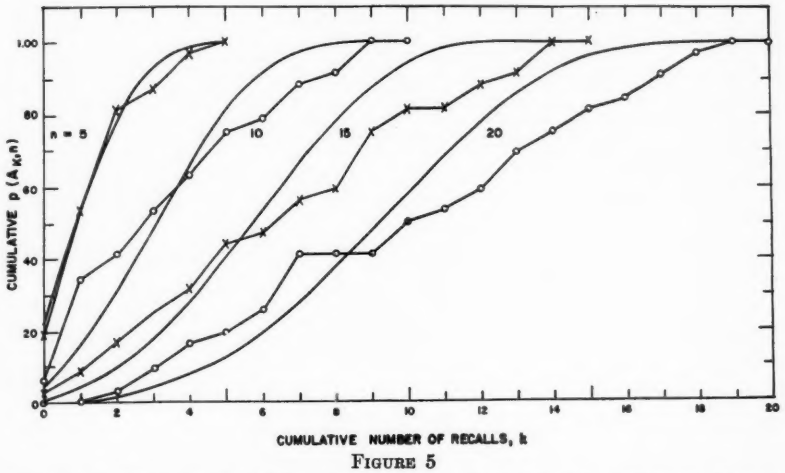


FIGURE 5

Comparison of Theoretical and Observed Distribution of Recalls on Four Different Trials in a Two-Parameter Case

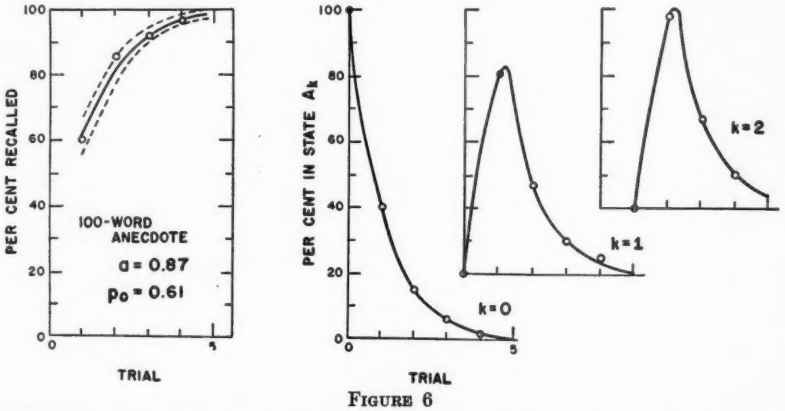


FIGURE 6

Comparison of Theoretical and Observed Values of ρ_n and $p(A_k, n)$ for a Two-Parameter Case

that τ_k and ρ_{n+1} approach the same asymptotic value, m. So we have the equation

$$\lim_{n\to\infty} \rho_{n+1} = m = \frac{a}{1-\alpha}. \tag{25}$$

Since $1 - \alpha \geq a$, m cannot exceed unity; and since both $a \geq 0$ and $1 - \alpha \geq 0$, m cannot be negative. In general, we are interested in cases where $m > p_0$, for if $p_0 > m$, we obtain forgetting rather than acquisition.

A set-theoretical rationalization for (24) runs as follows. On the presentation of the material a random sample of elements of measure p_0 is conditioned for every word. At the first recall a sample of measure $1 - \alpha$ is drawn. Of these elements, a portion of measure a is conditioned and the remainder, $1 - \alpha - a$, are extinguished. We add the conditioned elements as before, but now we must subtract the measure of the elements conditioned during presentation and extinguished during recall, i.e., $(1 - \alpha - a)\, p_0$. Thus we have

$$\begin{aligned}\tau_1 &= p_0 + a - ap_0 - (1 - \alpha - a)p_0 \\ &= m - (m - p_0)\alpha.\end{aligned}$$

At the second recall the same sampling procedure is repeated:

$$\begin{aligned}\tau_2 &= \tau_1 + a - a\tau_1 - (1 - \alpha - a)\tau_1 \\ &= a + \alpha\tau_1 = m - (m - p_0)\alpha^2.\end{aligned}$$

Continuing in this way generates the relation (24).

When (24) is substituted into (1), we obtain the appropriate difference equation, but its solution for the three-parameter case is hardly less cumbersome than (2). It would appear that the simplest way to work with these equations is to take advantage of our solution of the two-parameter case.

First, we introduce a new transitional probability, τ_k', such that

$$\begin{aligned}\tau_k' &= \tau_k/m \\ &= 1 - (1 - p_0/m)\alpha^k, \qquad \text{for } p_0 \leq m. \end{aligned} \tag{26}$$

This new variable is now the same as in the case of two parameters given in (18), with substitution of p_0/m for p_0 and α for $(1 - a)$. Therefore, from (2) and (20), we know that

$$\begin{aligned}\tau_0'\tau_1' \cdots \tau_{k-1}' \sum_{i=0}^{k} \frac{(1 - \tau_i')^n}{\prod\limits_{\substack{j=0 \\ j \neq i}}^{k} (\tau_i' - \tau_j')} &= \left(1 - \frac{p_0}{m}\right)^{n-k} \prod_{i=0}^{k-1} \frac{[1 - (1 - p_0/m)\alpha^i][1 - \alpha^{n-i}]}{1 - \alpha^{i+1}} \\ &= p'(A_k, n).\end{aligned} \tag{27}$$

When $m\ \tau_k'$ is substituted into (2), the factor m^k in the product in front of the summation cancels the factor m^k in the denominator under the summation. Thus we know that

$$p(A_k, n) = \tau_0'\tau_1' \cdots \tau_{k-1}' \sum_{i=0}^{k} \frac{(1 - \tau_i)^n}{\prod\limits_{\substack{j=0 \\ j \neq i}}^{k} (\tau_i' - \tau_j')}, \tag{28}$$

which is the same as $p'(A_k, n)$ in (27) except for the numerator under the summation. This numerator can be written

$$(1 - \tau_i)^n = [(1 - m) + m(1 - \tau_i')]^n$$

$$= (1 - m)^n + n(1 - m)^{n-1}m(1 - \tau_i')$$

$$+ \binom{n}{2}(1 - m)^{n-2}m^2(1 - \tau_i')^2 + \cdots m^n(1 - \tau_i')^n. \qquad (29)$$

Now we substitute this sequence for the numerator in (28) and sum term by term. When we consider the last term of this sequence we have

$$\tau_0'\tau_1' \cdots \tau_{k-1}' \sum_{i=0}^{k} \frac{m^n(1 - \tau_i')^n}{\prod_{\substack{j=0 \\ j \neq i}}^{k} (\tau_j' - \tau_i')},$$

which we know from (27) is equal to $m^n p'(A_k, n)$. The next to last term gives

$$\tau_0'\tau_1' \cdots \tau_{k-1}' \sum_{i=0}^{k} \frac{n(1 - m)^{n-1}(1 - \tau_i')^{n-1}}{\prod_{\substack{j=0 \\ j \neq i}}^{k} (\tau_j' - \tau_i')},$$

which we know from (27) is equal to $n(1 - m)m^{n-1}\, p'(A_k, n - 1)$. Proceeding in this manner brings us eventually to the case where $n < k$, and then we know the term is zero. Consequently, we can write

$$p(A_k, n) = m^n p'(A_k, n) + n(1 - m)m^{n-1}p'(A_k, n - 1) + \cdots$$

$$+ \binom{n}{n-k}(1 - m)^{n-k}m^k p'(A_k, k)$$

$$= \sum_{i=k}^{n} \binom{n}{i} m^i(1 - m)^{n-i}p'(A_k, i). \qquad (30)$$

When the asymptote is unity ($m = 1$), (29) and (30) reduce to the two-parameter case.

We recall that because of the way in which our probabilities were defined in (1), (30) can be written as

$$p(A_k, n) = \sum_{i=0}^{n} \binom{n}{i} m^i(1 - m)^{n-i}p'(A_k, i).$$

Now it is not difficult to find an expression for ρ_{n+1} in terms of the ρ'_i computed in the two-parameter case:

$$\begin{aligned}\rho_{n+1} &= \sum_{k=0}^{n} \tau_k p(A_k , n) \\ &= m \sum_{k=0}^{n} \tau'_k p(A_k , n) \\ &= m \sum_{k=0}^{n} \sum_{i=0}^{n} \binom{n}{i} m^i (1 - m)^{n-i} \tau'_k p'(A_k , i).\end{aligned}$$

If we invert the order of summation, we find that

$$\begin{aligned}\rho_{n+1} &= m \sum_{i=0}^{n} \binom{n}{i} m^i (1 - m)^{n-i} \sum_{k=0}^{n} \tau'_k p'(A_k , i) \\ &= m \sum_{i=0}^{n} \binom{n}{i} m^i (1 - m)^{n-i} \rho'_{i+1} . \qquad (31)\end{aligned}$$

The computation of ρ_{n+1} by this method involves two steps: first, the values of ρ'_{n+1} are calculated as in the two-parameter case with the substitution indicated in (26); second, these values of ρ'_{n+1} are weighted by the binomial expansion of $[m + (1 - m)]^n$ and then summed according to (31).

These computations can be abbreviated somewhat by using an approximation developed by Bush and Mosteller (personal communication). It is

$$\begin{aligned}\rho_{n+2} &= (2 + a + 2a\alpha)\rho_{n+1} - [a^2(1 - \alpha) + (1 + a)(1 + 2a\alpha)]\rho_n \\ &\quad + 3(1 - \alpha^2)(1 - a)\rho_n^2 - 2(1 - \alpha)(1 - \alpha^2)\rho_n^3 - 3(1 - \alpha^2)\rho_n\rho_{n+1} , \\ &\qquad (n \geq 1). \qquad (32)\end{aligned}$$

The approximation involves permitting the third moment of the distribution of the τ_k around ρ_n to go to zero on every trial.

The variance of r_{n+1} in the three-parameter case is

$$\text{Var}\,(r_{n+1}) = \frac{m}{aN} [\rho_{n+2} - (a + \alpha)\rho_{n+1}]. \qquad (33)$$

This expression for the variance of r_{n+1} follows directly from (7) and (10′). It is easily seen that (10′) can be written as follows:

$$\sum_{k=0}^{n} \tau_k^2 \, p(A_k , n) = \rho_{n+1} - N \text{ Var } (r_{n+1}). \qquad (34)$$

Substituting (34) in (7) and solving for Var (r_{n+1}) we find that

$$\text{Var}\,(r_{n+1}) = \frac{1}{N(1 - \alpha)} [\rho_{n+2} - (a + \alpha)\rho_{n+1}],$$

which, except for notation, is (33). The one-parameter and two-parameter variances (16) and (22) are special cases of this expression.

It is of interest to observe that when the limiting value, m, is substituted in (33) for ρ_{n+2} and ρ_{n+1}, the limiting variance is found to be binomial. That is,

$$\lim_{n\to\infty} \text{Var}\,(r_{n+1}) = \frac{m(1-m)}{N}.$$

This reflects the fact, established earlier in (5′), that as n grows very large the variance of the τ_k around m goes to zero.

In order to obtain a numerical example, we have taken the data from another subject in the experiment by Bruner and Zimmerman. Sixty-four monosyllabic English words were read aloud and the order of the words was scrambled before every presentation. A visual inspection of the data led us to choose an asymptote in the neighborhood of 0.7. This asymptote is drawn on the plot of the t_k in Figure 7 and on the plot of the r_n in Figure 8. Then we

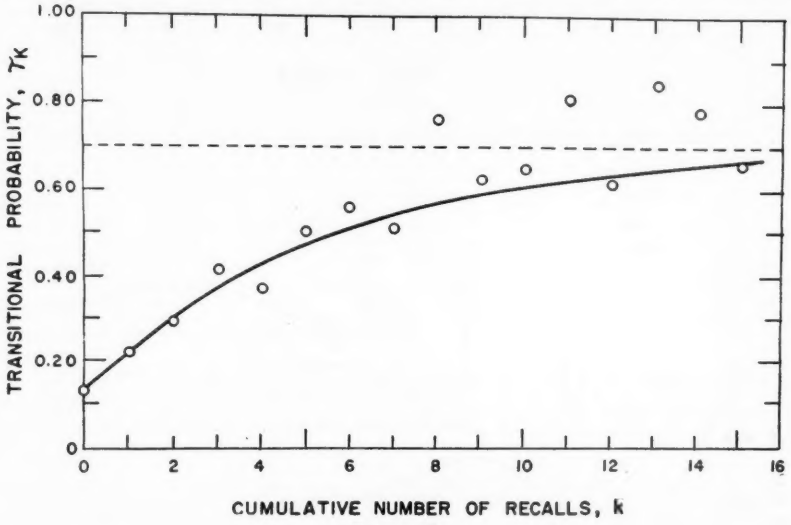


FIGURE 7

Transitional Probability of Recall, τ_k, as a Function of Number of Recalls in the Three-Parameter Case. Values of t_k are indicated by open circles. The curve fitted to the t_k is $\tau_k = 0.7 - 0.57\,(0.83)^k$.

estimated $p_0 = 0.13$ by considering all the trials on which words were in state A_0 and calculating p_0 as the weighted average of the $t_{0,n+1}$ for all those trials. Next we estimated the sampling parameter $\alpha = 0.83$. This was done by obtaining the estimates, t_k, for successive values of k; these estimates, together with (24), give us a set of equations estimating α. We used the weighted average of these estimates (ignoring negative values). Then we obtained $a = 0.12$ from the equation $a = m(1 - \alpha)$. We shall comment on the estimation problems later.

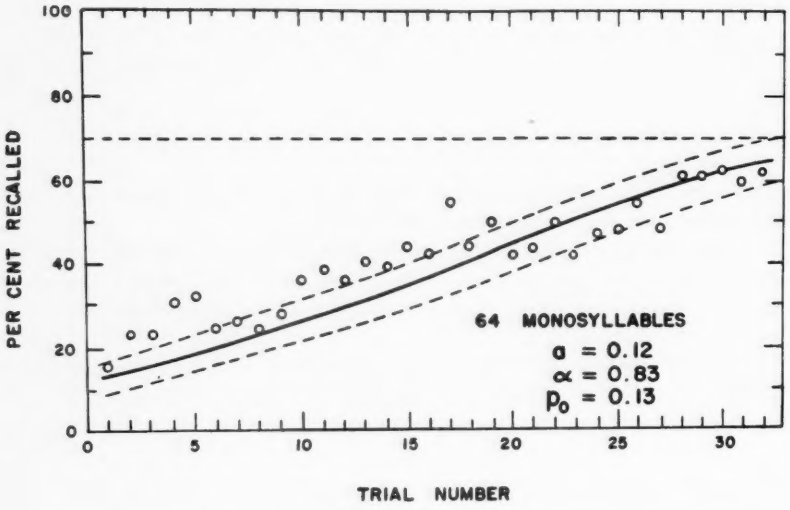


FIGURE 8

Comparison of Theoretical and Observed Values of ρ_n for Three-Parameter Case

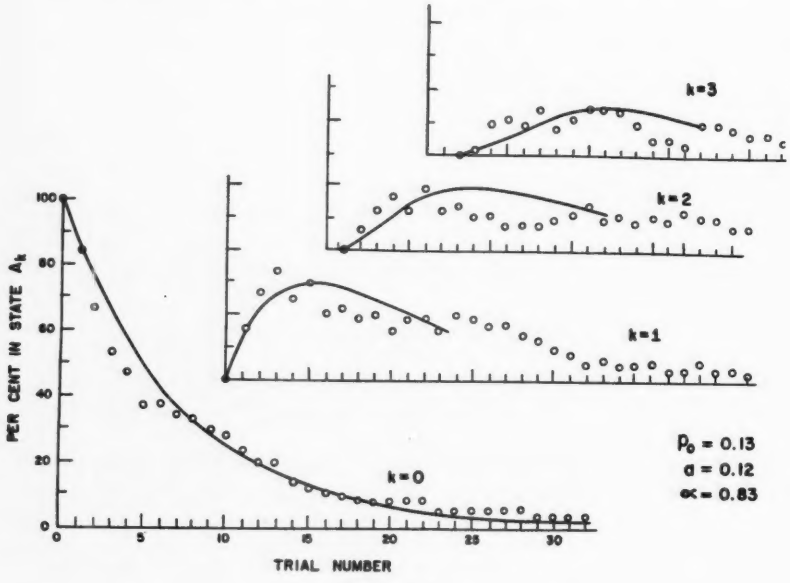


FIGURE 9

Comparison of Theoretical and Observed Values of $p(A_k , n)$ for Three-Parameter Case

When these parameter values were substituted into (24) we obtained the function for τ_k shown in Figure 7. When the values were substituted into (28) for $k = 1, 2, 3, 4$, we obtained the functions for $p(A_k, n)$ shown in Figure 9. When they were substituted into (31) we obtained the function, for ρ_n shown in Figure 8. In Figure 8 the dotted lines are drawn $\pm$ one standard deviation from ρ_n, as computed from (33).

A comparison of the values of ρ_n computed from (31) and from (32) is given for the first eighteen trials in Table 1. With this choice of parameters the Bush-Mosteller approximation seems highly satisfactory.

TABLE 1

Comparison of Exact and Approximate Values of ρ_n for First 18 Trials

Trial	Exact	Approximate	Trial	Exact	Approximate
1	.1300	.1300	10	.2663	.2655
2	.1426	.1426	11	.2837	.2827
3	.1559	.1559	12	.3014	.3000
4	.1700	.1700	13	.3191	.3174
5	.1847	.1846	14	.3369	.3347
6	.2000	.1999	15	.3546	.3520
7	.2159	.2157	16	.3722	.3692
8	.2323	.2319	17	.3896	.3862
9	.2491	.2486	18	.4067	.4030

Discussion

In the preceding pages we have made the explicit assumption that the several words being memorized simultaneously are independent, that memorizing one word does not affect the probability of recalling another word on the list. The assumption can be justified only by its mathematical convenience, because the data uniformly contradict it. The learner's introspective report is that groups of words go together to form associated clusters, and this impression is supported in the data by the fact that many pairs of words are recalled together or omitted together on successive trials. If the theory is used to describe the behavior of 50 rats, independence is a reasonable assumption. But when the theory describes the behavior of 50 words in a list that a single subject must learn, independence is not a reasonable assumption. It is important, therefore, to examine the consequences of introducing covariance.

The difference between the independent and the dependent versions of the theory can best be illustrated in terms of the set-theoretical interpretation of the two-parameter case. Imagine that we have a large ledger with 1000 pages. The presentation of the list is equivalent to writing each of the words

at random on 100 pages. Thus $p_0 = 100/1000 = 0.1$. Now we select a page at random. On this page we find written the words A, B, and C. These are responses on the first trial. The rule is that each of these words must be written on 50 pages selected at random. Thus $a = 50/1000 = 0.05$. With the independent model we would first select 50 pages at random and make sure that word A was written on all of them, then select 50 more pages independently for B, and 50 more for C. With a dependent model, however, we could simply make one selection of 50 pages at random and write all three words, A, B, and C, on the same sample of 50 pages. Then whenever A was recalled again it would be likely that B and C would also be recalled at the same time.

The probability that a word will be recalled depends upon the measure of the elements conditioned to it (the number of pages in the ledger on which it is inscribed) and does not depend upon what other words are written on the same pages. Therefore, the introduction of covariance in this way does not change the theoretical recall, ρ_{n+1}. The only effect is to increase the variance of the estimates of ρ_{n+1}. In other words, it is not surprising that the equations give a fair description of the recall scores even though no attention was paid to the probabilities of joint occurrences of pairs of words. Associative clustering should affect the variability, not the rate, of memorization.

The parameters a, p_0, and α obtained from the linear difference equation (6), are assumed to describe each word in the list. Thus data from different words may be combined to estimate the various τ_k. If the parameters vary from word to word, ρ_{n+1} is only an approximation of the mean probability of recall determined by averaging the recall probabilities of all the words. Similarly, the expressions given for ρ_{n+1} cannot be expected to describe the result of averaging several subjects' data together unless all subjects are known to have the same values of the parameters.

The general theory, of course, is not limited to linear restrictions of the form of (6). The data or the theory may force us to consider more complicated functions for τ_k. For all such cases the general solution (2) is applicable, though tedious to use, and will enable us to compute the necessary values of $p(A_k, n)$.

Once a descriptive model of this sort has been used to tease out the necessary parameters, the next step is to vary the experimental conditions and to observe the effects upon these parameters. In order to take this next step, however, we need efficient methods of estimating the parameters from the data. As yet we have found no satisfactory answers to the estimation problem.

There is a sizeable amount of computation involved in determining the functions $p(A_k, n)$ and ρ_n. If a poor choice of the parameters a, p_0, and α is made at the outset, it takes several hours to discover the fact. In the example in the preceding section, we estimated the parameters successively

and used different parts of the data for the different estimates. After ρ_n had been computed it seemed to us that our estimates of p_0 and m were both too low. Clearly, the method we have used to fit the theory to the data is not a particularly good one. We have considered least squares in order to use all of the data to estimate all parameters simultaneously. We convinced ourselves that the problem was beyond our abilities. Consequently, we must leave the estimation problem with the pious hope that it will appeal to someone with the mathematical competence to solve it.

Appendix A

Solution for $p(A_k, n)$ in the General Case

The solution of equation (1) with the boundary conditions we have enumerated has been obtained several times in the past (**4, 5**). We present below our own method of solution because the procedures involved may be of interest in other applications.

Equation (1) may be written explicitly as follows:

$$\begin{aligned}
(1 - \tau_0)p(A_0, n) &= p(A_0, n+1) \\
\tau_0 p(A_0, n) + (1 - \tau_1)p(A_1, n) &= p(A_1, n+1) \\
\tau_1 p(A_1, n) + (1 - \tau_2)p(A_2, n) &= p(A_2, n+1) \\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

This system of equations can be written in matrix notation as follows:

$$\begin{Bmatrix}
1-\tau_0 & 0 & 0 & 0 & \cdots \\
\tau_0 & 1-\tau_1 & 0 & 0 & \cdots \\
0 & \tau_1 & 1-\tau_2 & 0 & \cdots \\
0 & 0 & \tau_2 & 1-\tau_3 & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdots
\end{Bmatrix}
\begin{Bmatrix}
p(A_0, n) \\ p(A_1, n) \\ p(A_2, n) \\ p(A_3, n) \\ \cdot \\ \cdot \\ \cdot
\end{Bmatrix}
=
\begin{Bmatrix}
p(A_0, n+1) \\ p(A_1, n+1) \\ p(A_2, n+1) \\ p(A_3, n+1) \\ \cdot \\ \cdot \\ \cdot
\end{Bmatrix}$$

This infinite matrix of transitional probabilities we shall call T, and the infinite column vectors made up of the state probabilities on trial n and $n+1$ we shall call d_n and d_{n+1}. So we can write

$$Td_n = d_{n+1}.$$

The initial distribution of state probabilities, d_0, is the infinite column vector

$\{1, 0, 0, 0, \cdots\}$. The state probabilities on trial one are then given by

$$Td_0 = d_1 .$$

The state probabilities on trial two are given by

$$Td_1 = d_2 ,$$

so by substitution,

$$Td_1 = T(Td_0) = T^2 d_0 = d_2 .$$

Continuing this procedure gives the general relation

$$T^n d_0 = d_n .$$

Therefore, the problem of determining d_n can be equated to the problem of determining T^n.

Since T is a semi-matrix, we know that it can be expressed as

$$T = SDS^{-1},$$

where D is an infinite diagonal matrix with the same elements on its diagonal as are on the main diagonal of T (e.g., 2). The diagonal elements of S are arbitrary, so we let $S_{ii} = 1$. Now we can write

$$TS = SD$$

$$T\begin{Bmatrix} 1 & 0 & 0 & \cdot \\ S_{21} & 1 & 0 & \cdot \\ S_{31} & S_{32} & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix} = \begin{Bmatrix} 1 & 0 & 0 & \cdot \\ S_{21} & 1 & 0 & \cdot \\ S_{31} & S_{32} & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix} \begin{Bmatrix} 1-\tau_0 & 0 & 0 & \cdot \\ 0 & 1-\tau_1 & 0 & \cdot \\ 0 & 0 & 1-\tau_2 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix}.$$

Now it is a simple matter to solve for S_{ij} term by term. For example, to solve for S_{21} we construct (from row 2 and column 1) the equation

$$\tau_0 + (1 - \tau_1)S_{21} = S_{21}(1 - \tau_0),$$

which gives

$$S_{21} = \tau_0/(\tau_1 - \tau_0).$$

To solve for S_{31} , we use the equation

$$\begin{aligned} \tau_1 S_{21} + (1 - \tau_2)S_{31} &= S_{31}(1 - \tau_0) \\ S_{31} &= \tau_1 S_{21}/(\tau_2 - \tau_0) \\ &= \tau_0\tau_1/(\tau_1 - \tau_0)(\tau_2 - \tau_0). \end{aligned}$$

Proceeding in this manner gives the necessary elements of S, and we have

$$S = \begin{Bmatrix} 1 & 0 & 0 & 0 & \cdots \\ \frac{\tau_0}{(\tau_1 - \tau_0)} & 1 & 0 & 0 & \cdots \\ \frac{\tau_0\tau_1}{(\tau_1 - \tau_0)(\tau_2 - \tau_0)} & \frac{\tau_1}{(\tau_2 - \tau_1)} & 1 & 0 & \cdots \\ \frac{\tau_0\tau_1\tau_2}{(\tau_1 - \tau_0)(\tau_2 - \tau_0)(\tau_3 - \tau_0)} & \frac{\tau_1\tau_2}{(\tau_2 - \tau_1)(\tau_3 - \tau_1)} & \frac{\tau_2}{(\tau_3 - \tau_2)} & 1 & \cdots \\ . & . & . & . & \\ . & . & . & . & \\ . & . & . & . & \end{Bmatrix}$$

The elements of S^{-1} can be obtained term by term from the equation $SS^{-1} = 1$. For example, the element S'_{21} of S^{-1} is given by row two of S times column one of S^{-1} : $\tau_0/(\tau_1 - \tau_0) + S'_{21} = 0$. Continuing in this way we have

$$S^{-1} = \begin{Bmatrix} 1 & 0 & 0 & 0 & \cdots \\ \frac{\tau_0}{(\tau_0 - \tau_1)} & 1 & 0 & 0 & \cdots \\ \frac{\tau_0\tau_1}{(\tau_0 - \tau_2)(\tau_1 - \tau_2)} & \frac{\tau_1}{(\tau_1 - \tau_2)} & 1 & 0 & \cdots \\ \frac{\tau_0\tau_1\tau_2}{(\tau_0 - \tau_3)(\tau_1 - \tau_3)(\tau_2 - \tau_3)} & \frac{\tau_1\tau_2}{(\tau_1 - \tau_3)(\tau_2 - \tau_3)} & \frac{\tau_2}{(\tau_2 - \tau_3)} & 1 & \cdots \\ . & . & . & . & \\ . & . & . & . & \\ . & . & . & . & \end{Bmatrix}$$

These matrices permit a simple representation of the powers of the matrix T. Thus,

$$T^2 = (SDS^{-1})(SDS^{-1}) = SD(S^{-1}S)DS^{-1} = SD^2S^{-1},$$

and in general,

$$T^n = SD^nS^{-1}.$$

Since D is a diagonal matrix, D^n is obtained by taking the nth power of every

diagonal element. When this equation for T^n is multiplied through, we obtain

$$T^n = \begin{Bmatrix} (1-\tau_0)^n & 0 & 0 & \cdots \\ \tau_0\left[\dfrac{(1-\tau_0)^n}{(\tau_1-\tau_0)} + \dfrac{(1-\tau_1)^n}{(\tau_0-\tau_1)}\right] & (1-\tau_1)^n & 0 & \cdots \\ \tau_0\tau_1\left[\dfrac{(1-\tau_0)^n}{(\tau_1-\tau_0)(\tau_2-\tau_0)} + \dfrac{(1-\tau_1)^n}{(\tau_0-\tau_1)(\tau_2-\tau_1)} + \dfrac{(1-\tau_2)^n}{(\tau_0-\tau_2)(\tau_1-\tau_2)}\right] & \tau_1\left[\dfrac{(1-\tau_1)^n}{(\tau_2-\tau_1)} + \dfrac{(1-\tau_2)^n}{(\tau_1-\tau_2)}\right] & (1-\tau_2)^n & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{Bmatrix}$$

Since $T^n d_0$ involves only the first column of T^n, it is not actually necessary to obtain more than the first columns of S^{-1} and of T^n. We have presented the complete solution here, however. It can be seen from inspection of the first column of T^n that (2) is the general solution:

$$\begin{aligned} p(A_0, n) &= (1-\tau_0)^n, && \text{for } k = 0, \\ p(A_k, n) &= \tau_0\tau_1 \cdots \tau_{k-1} \sum_{i=0}^{k} \frac{(1-\tau_i)^n}{\prod_{\substack{j=0 \\ j\neq i}}^{k} (\tau_j - \tau_i)}, && \text{for } k > 0. \end{aligned} \tag{2}$$

This general method of solution can be used for the special cases considered in this paper, with the substitution of the appropriate values for τ_k.

Appendix B

Recursive Expression for ρ_{n+1} in Two-Parameter Case

From (20) we obtain the recursive relation

$$p(A_{k+1}, n+1) = \frac{[1-(1-p_0)(1-a)^k][1-(1-a)^{n+1}]}{1-(1-a)^{k+1}}\, p(A_k, n).$$

Rearranging and summing, we have

$$\sum_{k=0}^{n}\left[\frac{1-(1-a)^{k+1}}{1-(1-a)^{n+1}}\, p(A_{k+1}, n+1)\right] = \sum_{k=0}^{n}[1-(1-p_0)(1-a)^k]p(A_k, n).$$

The right side of this equation is, from (5) and (18), ρ_{n+1}. The left side can be rewritten

$$\sum_{k=1}^{n+1} \left[\frac{1 - (1 - a)^k}{1 - (1 - a)^{n+1}} p(A_k , n + 1) \right] = \rho_{n+1} ,$$

which becomes on trial n (with $n \geq 1$),

$$\sum_{k=1}^{n} \left[\frac{1 - (1 - a)^k}{1 - (1 - a)^n} p(A_k , n) \right] = \rho_n .$$

We now have, by adding and subtracting $p(A_0 , n)$,

$$\frac{1}{1 - (1 - a)^n} \left[\sum_{k=0}^{n} p(A_k , n) - \sum_{k=0}^{n} (1 - a)^k p(A_k , n) \right] = \rho_n ,$$

$$1 - \sum_{k=0}^{n} (1 - a)^k p(A_k , n) = [1 - (1 - a)^n]\rho_n .$$

Now we know that

$$\rho_{n+1} = 1 - (1 - p_0) \sum_{k=0}^{n} (1 - a)^k p(A_k , n),$$

and so we obtain

$$\rho_{n+1} = 1 - (1 - p_0)\{1 - [1 - (1 - a)^n]\rho_n\}.$$

Rearranging terms gives

$$\rho_{n+1} = p_0 + (1 - p_0)[1 - (1 - a)^n]\rho_n , \tag{21}$$

which is the desired result.

From this result (15) is obtained directly by equating p_0 and a.

Appendix C

List of Symbols and Their Meanings

a	parameter.
A_k	state that a word is in after being recalled k times.
α	parameter.
d_n	infinite column vector, having $p(A_k , n)$ as its elements.
D	infinite diagonal matrix similar to T.
k	number of times a word has been recalled.
m	asymptotic value of τ_k and ρ_n.
n	number of trial.
N	total number of test words to be learned.
$N_{k,n}$	number of words in state A_k on trial n.
p_0	probability of recalling a word in state A_0.

$p(A_k, n)$	probability that a word will be in state A_k on trial n.
r_n	observed recall score on trial n; estimate of ρ_n.
ρ_n	probability of recall on trial n.
S_{ij}	elements of S.
S'_{ij}	elements of S^{-1}.
S	infinite matrix used to transform T into a similar diagonal matrix.
t_k	estimate of τ_k.
$t_{k,n}$	observed fraction of words in state A_k that are recalled on trial n.
τ_k	probability of recalling a word in state A_k.
T	infinite matrix of transition probabilities τ_k.
Var (r_n)	variance of the estimate of ρ_n.
$X_{i,k,n+1}$	random variable equal to 1 or 0.

REFERENCES

1. Bush, R. R., and Mosteller, Frederick. A linear operator model for learning. (Paper presented to the Institute for Mathematical Statistics, Boston, December 27, 1951.)
2. Cooke, R. G. Infinite matrices and sequence spaces. London: MacMillan, 1950.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Feller, W. On the theory of stochastic processes with particular reference to applications. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1949, 403–432.
5. Woodbury, M. A. On a probability distribution. *Ann. math. Statist.*, 1949, **20**, 311–313.

Manuscript received 3/11/52

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

NOTE ON THE SCALING OF RATINGS OR RANKINGS WHEN THE NUMBERS PER SUBJECT ARE UNEQUAL

EDWARD E. CURETON

UNIVERSITY OF TENNESSEE

The average rating (or normalized rank) of a person rated by a larger number of judges will in general be closer to the group average than will the average rating of a person rated by a smaller number of judges, as a result of rating unreliability and regression. This note presents a technique for correcting that bias.

In criterion development and merit evaluation it is common practice to have each subject rated or ranked by as many judges as consider themselves competent to do so. The investigator ordinarily starts by preparing a complete alphabetical list of the subjects. A copy of this list is given to each judge, who first crosses off the names of those he does not know well, and then rates or ranks the remainder. A comparable situation is encountered in selection when a rating blank is substituted for the usual letter of reference. In either case the numbers of ratings or rankings for different subjects are unequal, and any degree of overlap from zero to 100 per cent may be found among the subjects rated by any one judge and those rated by any other. In this situation, differences in rating standard from judge to judge must be considered a part of the error variance.

A method for estimating the average reliability of the single rating has been given recently by Ebel.* His method may be outlined as follows:

Let X be a rating (or normalized rank),
N the number of subjects in the total group,
n_i the number of ratings of the ith subject,
M the mean of all ratings,
M_i the mean of the ratings of the ith subject,
$\sum$ a summation from 1 to N,
S a summation from 1 to n_i,
s_e^2 the within-persons variance of X (the error variance),
s_p^2 the between-persons variance of X, and
k a special form of average of the N values of n_i.

*Ebel, Robert L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407–424.

Then

$$M = \frac{\sum SX}{\sum n_i}, \tag{1}$$

$$s_e^2 = \frac{\sum SX^2 - M \sum SX}{\sum n_i - N}, \tag{2}$$

$$M_i = \frac{SX_i}{n_i}, \tag{3}$$

$$s_p^2 = \frac{\sum (M_i SX_i) - M \sum SX}{N - 1}, \tag{4}$$

$$k = \frac{(\sum n_i)^2 - \sum n_i^2}{(N - 1) \sum n_i}, \tag{5}$$

and

$$r_1 = \frac{s_p^2 - s_e^2}{s_p^2 + (k - 1)s_e^2}. \tag{6}$$

The values of $\sum SX$, $\sum SX^2$, and $\sum n_i$ can be obtained directly from the entire set of ratings or normalized rankings. The ratings of each subject are then counted and summed, yielding n_i, SX_i, and M_i, and from these values we can compute $\sum n_i$, $\sum n_i^2$, and $\sum (M_i SX_i)$. The value of $\sum n_i$ obtained at this step serves as a partial check. In computing these values we cannot use any subject who has been rated by only one judge. r_1 is the average reliability of the single rating.

If we wish to compare the average ratings of different subjects, the situation is complicated by the fact that the error variance is greater for a subject rated by a smaller number of judges than for a subject rated by a larger number. The extreme average ratings at both ends of the scale will tend to go in part to subjects rated by small numbers of judges rather than simply to those rated high and low. This bias is systematic, and may be removed by computing the estimated "true" rating for each subject, $X_{i\infty}$, and using it instead of the average raw rating, M_i. To do this we compute the value of r_i for each subject by the Spearman-Brown formula,

$$r_i = \frac{n_i r_1}{1 + (n_i - 1)r_1}. \tag{7}$$

Then

$$X_{i\infty} = M(1 - r_i) + r_i M_i. \tag{8}$$

This last equation gives an unbiased estimate of the "true" rating.

If the original group contains persons rated by only one rater, formula (8) may be applied to them also. For such persons, r_i is r_1 and M_i is X_i.

The value of r_1 is still estimated, of course, from those cases for which two or more ratings are available. If persons having only one rating are included in the final group, a new value of M, including the scores of these persons, should be computed by (1) for substitution in (8).

Manuscript received 4/30/52

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

MULTIDIMENSIONAL SCALING: I. THEORY AND METHOD*

WARREN S. TORGERSON

SOCIAL SCIENCE RESEARCH COUNCIL

Multidimensional scaling can be considered as involving three basic steps. In the first step, a scale of comparative distances between all pairs of stimuli is obtained. This scale is analogous to the scale of stimuli obtained in the traditional paired comparisons methods. In this scale, however, instead of locating each stimulus-object on a given continuum, the distances between each pair of stimuli are located on a distance continuum. As in paired comparisons, the procedures for obtaining a scale of comparative distances leave the true zero point undetermined. Hence, a comparative distance is not a distance in the usual sense of the term, but is a distance minus an unknown constant. The second step involves estimating this unknown constant. When the unknown constant is obtained, the comparative distances can be converted into absolute distances. In the third step, the dimensionality of the psychological space necessary to account for these absolute distances is determined, and the projections of stimuli on axes of this space are obtained. A set of analytical procedures was developed for each of the three steps given above, including a least-squares solution for obtaining comparative distances by the complete method of triads, two practical methods for estimating the additive constant, and an extension of Young and Householder's Euclidean model to include procedures for obtaining the projections of stimuli on axes from fallible absolute distances.

Introduction

The traditional methods of psychophysical scaling presuppose knowledge of the dimensions of the area being investigated. The methods require judgments along a particular defined dimension, i.e., *A* is brighter, twice as loud, more conservative, or heavier than *B*. The observer, of course, must know what the experimenter means by brightness, loudness, etc. In many stimulus domains, however, the dimensions themselves, or even the number of relevant dimensions, are not known. What might appear intuitively to be a single dimension may in fact be a complex of several. Some of the intuitively given dimensions may not be necessary—it may be that they can be accounted for by linear combinations of others. Other dimensions of importance may be completely overlooked. In such areas the traditional approach is inadequate.

Richardson, in 1938 (3; see also Gulliksen, 1) proposed a model for multidimensional scaling that would appear to be applicable to a number of

*This study was carried out while the author was an Educational Testing Service Psychometric Fellow at Princeton University. The author expresses his appreciation to his thesis adviser, Dr. H. Gulliksen, for his guidance throughout the study and to Dr. B. F. Green, Jr., for valuable assistance on several of the derivations.

these more complex areas. This model differs from the traditional scaling methods in two important respects. First, it does not require judgments along a given dimension, but utilizes, instead, judgments of similarity between the stimuli. Second, the dimensionality, as well as the scale values, of the stimuli is determined from the data themselves.

Multidimensional scaling may perhaps best be considered as involving three basic steps. In the first step, a scale of comparative distances between all pairs of stimuli is obtained. The second step involves estimating an additive constant and using this estimate to convert the comparative distances into absolute distances. In the third step, the dimensionality of the psychological space necessary to account for these absolute distances is determined, and the projections of the stimuli on axes of this space are obtained.

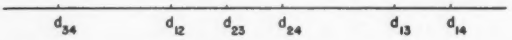
The Scale of Comparative Distances

The scale of comparative distances obtained in the multidimensional methods is analogous to the one-dimensional scale of stimulus-objects obtained in the traditional paired comparison type methods.

In the one-dimensional methods, the obtained scale locates the stimulus-objects with respect to one another on the given continuum. For example, given four stimulus-objects designated S_1 , S_2 , S_3 , and S_4 , the one-dimensional procedure might yield the following scale:

In this scale, the locations of the stimuli relative to one another only are determined from the data. The zero point of the scale is arbitrary. While the usual procedure is to locate the zero point so as to coincide with the stimulus having the lowest scale value, any other finite location on the continuum would serve equally well.

In the analogous scale of comparative distances obtained in the multidimensional procedures, the element, instead of being a stimulus-object, is a distance between two stimuli. Thus, given the same four stimulus-objects, the scale of comparative distances locates, with respect to one another on a distance continuum, the six inter-stimulus distances, d_{12} , d_{13} , d_{14} , d_{23} , d_{24} , and d_{34} :



The locations of the inter-stimulus distances relative to one another only are determined from the data. The zero point is again arbitrarily selected. It is important to note, however, that a comparative distance is not a "distance" in the usual sense of the term, but is a distance minus an unknown constant. In order to obtain absolute distances between stimuli, it is necessary to

estimate this constant. This is equivalent to estimating the true zero point of the scale of comparative distances. Thus, a comparative distance h_{jk} plus an unknown additive constant C gives the corresponding absolute distance d_{jk} :

$$h_{jk} + C = d_{jk} .$$

The Additive Constant for Converting Comparative Distances into Absolute Distances

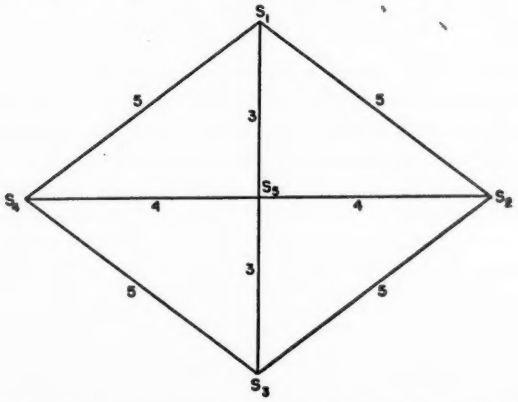
In estimating the additive constant, it is assumed that that value which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality is the value wanted. Consider, for example, five points having the following comparative interpoint distances h_{jk} ($j, k = 1, 2, \ldots, 5; j \neq k$):

$$h_{12} = 1, \quad h_{14} = 1, \quad h_{23} = 1, \quad h_{25} = 0, \quad h_{35} = -1,$$
$$h_{13} = 2, \quad h_{15} = -1, \quad h_{24} = 4, \quad h_{34} = 1, \quad h_{45} = 0.$$

With these comparative distances the value of the additive constant which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality is 4. If we add 4 to each of the comparative distances to convert them into absolute distances we obtain

$$d_{12} = 5, \quad d_{14} = 5, \quad d_{25} = 4, \quad d_{15} = 3, \quad d_{35} = 3,$$
$$d_{13} = 6, \quad d_{24} = 8, \quad d_{34} = 5, \quad d_{23} = 5, \quad d_{45} = 4.$$

The five stimuli can be plotted in a two-dimensional space:



Note that for any smaller value of the additive constant the points do not exist in a real Euclidean space. For example, if 1, 2, or 3 is added, then $d_{45} + d_{25} < d_{24}$, an impossible relationship in real Euclidean space. Also, for any larger value of the additive constant, the points lie in a real space of dimensionality greater than two.

Determination of the Dimensionality of the Psychological Space and the Projections of the Stimuli on Axes of the Space from the Absolute Distances Between the Stimuli

Young and Householder (5) have given a method for determining whether a set of absolute interpoint distances can be considered to be the distances between points lying in a real Euclidean space. They also have given, provided that the distances can so be considered, methods for determining the dimensionality of the space, and the projections of the points on a set of orthogonal axes of the space. Their theorems involve two basic matrices, B_i and F.

If we let

i, j, and k be alternate subscripts for n points ($i, j, k = 1, 2, \ldots, n$) and d_{ij} , d_{ik} , and d_{jk} be the distances between the points, then B_i is an $(n - 1) \times (n - 1)$ symmetric matrix with elements

$$b_{jk} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2). \tag{1}$$

The element b_{jk} may be considered to be the scalar product of vectors from point i to points j and k. This follows directly from the cosine law. That is, given the three points i, j, and k,

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik} \cos \theta_{jik} ,$$

which rearranged becomes

$$d_{ij}d_{ik} \cos \theta_{jik} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2). \tag{3}$$

From Equations 1 and 3, it is seen that $b_{jk} = d_{ij}d_{ik} \cos \theta_{jik}$, the scalar product of vectors from point i to points j and k. Matrix B_i is thus a matrix of scalar products of vectors with origin at point i. There are, of course, n possible B_i matrices, since i may assume any value from 1 to n.

Matrix F is an $(n + 1) \times (n + 1)$ symmetric matrix of squares of interpoint distances bordered by a row and column of ones as follows:

$$F = \begin{matrix} 0 & d_{12}^2 & \cdots & d_{1k}^2 & \cdots & d_{1n}^2 & 1 \\ d_{21}^2 & 0 & & & & d_{2n}^2 & 1 \\ \cdot & & & & & \cdot & \cdot \\ \cdot & & & & & \cdot & \cdot \\ \cdot & & & & & \cdot & \cdot \\ d_{j1}^2 & & & & & d_{jn}^2 & 1 \\ \cdot & & & & & \cdot & \cdot \\ \cdot & & & & & \cdot & \cdot \\ \cdot & & & & & \cdot & \cdot \\ d_{n1}^2 & d_{n2}^2 & \cdots & d_{nk}^2 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & \cdots & 1 & 0 \end{matrix} \tag{4}$$

Young and Householder have shown that:

1. If any matrix B_i is positive semidefinite, the distances may be considered to be the distances between points lying in a real Euclidean space.
2. The rank of any positive semidefinite matrix B_i is equal to the dimensionality of the set of points.
3. The rank of matrix F is two greater than the dimensionality of the set of points.
4. Any positive semidefinite matrix B_i may be factored to obtain a matrix A_i such that

$$B_i = A_i A_i' . \tag{5}$$

If the rank of B_i is r, where $r \leq (n - 1)$, then matrix A_i is an $(n - 1) \times r$ matrix of projections of points on r orthogonal axes with origin at the ith point of the r-dimensional, real Euclidean space.

It is interesting to note that except for Richardson's original experiment (only an abstract of which has been published) only one person, Klingberg, (2) has used the model. It may well be that one of the reasons for the lack of experimental investigation in this area is that no clear statement of analytical procedure has been published. The problem of precisely how to proceed in obtaining comparative distances from proportions of judgments has not been adequately answered for either Richardson's method of triadic combinations or Klingberg's method of multidimensional rank order. While the analogy between the logic of paired comparisons and both of these methods is clear, the procedures cannot be directly applied in obtaining an efficient estimate. The least-squares solution for paired comparisons scales cannot be used because the analogous proportion matrix contains a rather large number of vacant cells—neither multidimensional method obtains judgments of the differences in distance between *all* possible pairs of distances, but only between pairs having one stimulus in common. Furthermore, in reducing the matrix of distance-differences between pairs to a scale of comparative distances, one is almost overwhelmed by the great number of possible modes of attack—each likely to give a somewhat different answer due to error in the observed data.

The problem of how to obtain a best estimate of the unknown additive constant has not been answered. The method used by Klingberg is quite tedious (it involved obtaining two tenth-order polynomials from the fifth-order minors and then solving for the unknowns) and does not insure that the answer obtained is a best estimate, or that it even approximates the value desired.

Similarly, while Young and Householder give adequate procedures for obtaining projections of points on axes from distances when the data are infallible, a number of difficulties arise when fallible data are employed.

The purpose of the present paper is to present a set of analytical procedures for multidimensional scaling, including, as far as possible, routine procedures for obtaining comparative distances, for estimating the additive constant, and for obtaining projections of stimuli on axes when fallible absolute distances are given. We shall first consider the complete method of triads for obtaining comparative distances between the stimuli. Following this, the problem of obtaining projections of stimuli on axes from fallible absolute distances will be discussed. Finally, we shall consider various methods for estimating the unknown additive constant.

The Complete Method of Triads for Obtaining Comparative Distances Between Stimuli

The stimuli are presented to the subject in triads. The judgment required of the subject is of the form: "Stimulus k is more similar to stimulus j than to stimulus i." With n stimuli, there are $n(n-1)(n-2)/6$ triads. In each triad, each stimulus is compared with each other pair, making a total of $n(n-1)(n-2)/2$ judgments for each subject. From these judgments we obtain the proportion of times any stimulus k is judged more similar to stimulus j than to i.

These proportions can be arranged in the n matrices ${}_kP_{ij}$ where k, i, and j are alternate subscripts for the stimuli. k gives the number of the matrix, i is a row index, and j is a column index. The element ${}_kp_{ij}$ is the proportion of times stimulus k is judged closer to stimulus j than to i. The matrices ${}_kP_{ij}$ have vacant cells in the principal diagonal, and in the kth row and column.* The matrices are such that the sum of symmetric elements is unity—e.g., ${}_kp_{gh} + {}_kp_{hg} = 1$. For example, given four stimuli, 1, 2, 3, and 4, there are four ${}_kP_{ij}$ matrices. The second matrix ($k = 2$) is illustrated below:

${}_2P_{ij}$

	1	2	3	4
1		·	${}_2p_{13}$	${}_2p_{14}$
2	·		·	·
3	${}_2p_{31}$	·		${}_2p_{34}$
4	${}_2p_{41}$	·	${}_2p_{43}$	

The first problem is to transform the proportions ${}_kp_{ij}$ into differences in distances ${}_kx_{ij}$. We shall assume that the proportion of times stimulus k is judged closer to stimulus j than to i is a function of the difference in the

*It might be noted that the elements in the kth row and column *could* be obtained experimentally. However, since the method would ordinarily be used in connection with supraliminal distances, the experimentally determined proportions would be either .00 or 1.00. As in paired comparisons, proportions of .00 and 1.00 cannot be utilized.

distances, $d_{ki} - d_{kj} = {}_kx_{ij}$, the function being

$$ {}_kp_{ij} = \int_{-\infty}^{{}_kx_{ij}} \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}x^2} \, dx. \tag{6} $$

${}_kx_{ij}$ is thus $\sqrt{2}$ times the deviate of the unit normal curve measured in σ units from the mean. This is analogous to Thurstone's Case V in paired comparisons (4) and is the same assumption used previously by Richardson (3).

Making this transformation we obtain the n matrices ${}_kX_{ij}$. These matrices are skew symmetric (${}_kx_{gh} + {}_kx_{hg} = 0$), have zero diagonal elements, and have vacant cells in the kth row and column.

We have $n(n-1)(n-2)/2$ independent observations of differences in distances ${}_kx_{ij}$ from which we wish to determine $n(n-1)/2$ comparative distances h_{ik}. Since $\frac{1}{2}n(n-1) - 1$ differences in distances are sufficient to determine a matrix of comparative distances, it is apparent that the data are considerably overdetermined. There are, of course, a large number of sets of $\frac{1}{2}n(n-1) - 1$ differences in distances which could be used. Also, there are many different ways of obtaining the comparative distances from each set. With fallible data, the matrices could be expected to differ somewhat from each other.

The first problem, then, is to find a best estimate, in a least-squares sense, of the matrix of comparative distances h_{ik} in terms of the available data.

The element ${}_kx_{ij} = d_{ik} - d_{kj} + {}_ke_{ij}$, where d_{ik} and d_{kj} are *absolute* distances between stimuli k and i, and k and j, respectively, and ${}_ke_{ij}$ is an error.* It would seem that we want that set of interpoint distances which minimizes the sum of squares of the errors ${}_ke_{ij}$. For a least-squares solution, then, we wish to select the distances to minimize the following function:

$$ 2F = \sum_{k}^{n} \sum_{\substack{j \\ j \neq k}}^{n} \sum_{\substack{i \\ i \neq j \\ i \neq k}}^{n} [{}_kx_{ij} - (d_{ik} - d_{kj})]^2. \tag{7} $$

If we define a set of matrices ${}_kE_{ij}$ with elements $({}_kx_{ij} - d_{ik} + d_{kj})$, it is seen that $2F$ is equal to the sum of squares of elements of the matrices ${}_kE_{ij}$.

Let g and h correspond to two particular stimuli with $d_{gh} = d_{hg}$, the distance between them. The term d_{gh} (or d_{hg}) occurs only in the error matrices ${}_gE_{ij}$ and ${}_hE_{ij}$ as follows:

in ${}_gE_{ij}$ the hth column contains the elements ${}_gx_{ih} - d_{ig} + d_{gh}$,
the hth row contains the elements ${}_gx_{hj} - d_{hg} + d_{gj}$;
in ${}_hE_{ij}$ the gth column contains the elements ${}_hx_{ig} - d_{ih} + d_{hg}$,
the gth row contains the elements ${}_hx_{gj} - d_{gh} + d_{hj}$.

*${}_kx_{ij}$ is also equal to the difference between the comparative distances, since the difference in absolute distances $d_{ik} - d_{kj}$ is identical with the difference in comparative distances $(d_{ik} - C) - (d_{kj} - C)$.

To minimize $2F$, we first take the derivative of F with respect to d_{gh}. We shall designate this derivative as F'. It is apparent that the derivatives of all terms of F except those containing the element d_{gh} vanish. Therefore,

$$F' = \sum_{\substack{i \\ i \neq g,h}}^{n} ({}_{g}x_{ih} - d_{ig} + d_{gh}) - \sum_{\substack{j \\ j \neq g,h}}^{n} ({}_{g}x_{hj} - d_{hg} + d_{gj})$$

$$+ \sum_{\substack{i \\ i \neq g,h}}^{n} ({}_{h}x_{ig} - d_{ih} + d_{hg}) - \sum_{\substack{j \\ j \neq g,h}}^{n} ({}_{h}x_{gj} - d_{gh} + d_{hj}). \tag{8}$$

But, since matrices ${}_{g}E_{ij}$ and ${}_{h}E_{ij}$ are skew symmetric,

$$\sum_{\substack{i \\ i \neq g,h}} ({}_{g}x_{ih} - d_{ig} + d_{gh}) = - \sum_{\substack{j \\ j \neq g,h}} ({}_{g}x_{hj} - d_{hg} + d_{gj}), \tag{9}$$

and

$$\sum_{\substack{i \\ i \neq g,h}} ({}_{h}x_{ig} - d_{ih} + d_{hg}) = - \sum_{\substack{j \\ j \neq g,h}} ({}_{h}x_{gj} - d_{gh} + d_{hj}). \tag{10}$$

Therefore, we may write

$$F' = 2 \sum_{\substack{i \\ i \neq g,h}} ({}_{g}x_{ih} - d_{ig} + d_{gh}) + 2 \sum_{\substack{i \\ i \neq g,h}} ({}_{h}x_{ig} - d_{ih} + d_{hg}). \tag{11}$$

Setting F' equal to zero, and summing over each term, we find

$$\sum_{\substack{i \\ i \neq g,h}} {}_{g}x_{ih} - \sum_{\substack{i \\ i \neq g,h}} d_{ig} + (n-2)d_{gh} + \sum_{\substack{i \\ i \neq g,h}} {}_{h}x_{ig} - \sum_{\substack{i \\ i \neq g,h}} d_{ih}$$

$$+ (n-2)d_{hg} = 0. \tag{12}$$

Remembering that $d_{gg} = d_{hh} = 0$, and that the diagonals and kth row and column of all ${}_{k}X_{ij}$ matrices are vacant, we can write:

$$\sum_{i} {}_{g}x_{ih} - \sum_{\substack{i \\ i \neq h}} d_{ig} + (n-2)d_{gh} + \sum_{i} {}_{h}x_{ig} - \sum_{\substack{i \\ i \neq g}} d_{ih}$$

$$+ (n-2)d_{hg} = 0. \tag{13}$$

Subtracting d_{gh} from $- \sum_{\substack{i \\ i \neq h}} d_{ig}$ and from $- \sum_{\substack{i \\ i \neq g}} d_{ih}$, adding d_{gh} to $(n-2)d_{gh}$ and $(n-2)d_{hg}$, and remembering that $d_{gh} = d_{hg}$, we have

$$\sum_{i} {}_{g}x_{ih} - \sum_{i} d_{ig} + (n-1)d_{gh} + \sum_{i} {}_{h}x_{ig} - \sum_{i} d_{ih}$$

$$+ (n-1)d_{hg} = 0. \tag{14}$$

Summing over g, $g \neq h$, we have

$$\sum_{\substack{g \\ g \neq h}} \sum_{i} {}_{g}x_{ih} - \sum_{\substack{g \\ g \neq h}} \sum_{i} d_{ig} + (n-1) \sum_{\substack{g \\ g \neq h}} d_{gh} + \sum_{\substack{g \\ g \neq h}} \sum_{i} {}_{h}x_{ig}$$

$$- (n-1) \sum_{i} d_{ih} + (n-1) \sum_{\substack{g \\ g \neq h}} d_{hg} = 0. \tag{15}$$

But

$$\sum_{\substack{g \\ g \neq h}} \sum_{i} {}_{h}x_{ig} = 0, \tag{16}$$

and, since $d_{hh} = 0$,

$$\sum_{i} d_{ih} = \sum_{\substack{g \\ g \neq h}} d_{hg} ; \tag{17}$$

therefore,

$$\sum_{\substack{g \\ g \neq h}} \sum_{i} {}_{g}x_{ih} - \sum_{\substack{g \\ g \neq h}} \sum_{i} d_{ig} + (n-1) \sum_{g} d_{gh} = 0. \tag{18}$$

Subtracting $\sum_{i} d_{ih}$ from $\sum_{\substack{g \\ g \neq h}} \sum_{i} d_{ig}$ and adding $\sum_{g} d_{gh}$ to $(n-1) \sum_{g} d_{gh}$, we see, from Equation (17), that

$$\sum_{\substack{g \\ g \neq h}} \sum_{i} {}_{g}x_{ih} - \sum_{g} \sum_{i} d_{ig} + n \sum_{g} d_{gh} = 0. \tag{19}$$

Rearranging, dividing by $n(n-1)$, and remembering that cells ${}_{g}x_{ig}$ are vacant, we find

$$\frac{1}{n(n-1)} \sum_{g} \sum_{i} {}_{g}x_{ih} = \frac{1}{n(n-1)} \sum_{g} \sum_{i} d_{ig} - \frac{1}{(n-1)} \sum_{g} d_{gh} . \tag{20}$$

Also, if we divide Equation (14) by $2(n-1)$ and rearrange, we obtain

$$\frac{1}{2(n-1)} \left[\sum_{i} d_{ig} + \sum_{i} d_{ih} \right] - d_{gh} = \frac{1}{2(n-1)} \left[\sum_{i} {}_{g}x_{ih} + \sum_{i} {}_{h}x_{ig} \right]. \tag{21}$$

It will be convenient to define the averages in Equations (20) and (21) as follows:

$$d_{.h} \equiv \frac{1}{(n-1)} \sum_{i} d_{ih} = \frac{1}{(n-1)} \sum_{g} d_{gh} , \tag{22}$$

$$d_{.g} \equiv \frac{1}{(n-1)} \sum_{i} d_{ig} , \tag{23}$$

$$d_{..} \equiv \frac{1}{n(n-1)} \sum_{i} \sum_{g} d_{ig} , \tag{24}$$

$${}_{g}x_{.h} \equiv \frac{1}{(n-1)} \sum_{i} {}_{g}x_{ih} , \tag{25}$$

$${}_{h}x_{.g} \equiv \frac{1}{(n-1)} \sum_{i} {}_{h}x_{ig} , \tag{26}$$

$${}_{.}x_{gh} \equiv \frac{1}{n(n-1)} \sum_{i} \sum_{g} {}_{g}x_{ih} . \tag{27}$$

After substitutions have been made for the appropriate terms, Equation (21) becomes,

$$\tfrac{1}{2}(d_{.g} + d_{.h}) - d_{gh} = \tfrac{1}{2}({}_{g}x_{.h} + {}_{h}x_{.g}), \tag{28}$$

and Equation (20) becomes

$$.x_{.h} = d_{..} - d_{.h}\,, \tag{29}$$

and, when $h = g$,

$$.x_{.g} = d_{..} - d_{.g}\,.$$

Substituting for $d_{.g}$ and $d_{.h}$ in Equation (28), we have

$$\tfrac{1}{2}(d_{..} - .x_{.g} + d_{..} - .x_{.h}) - a_{gh} = \tfrac{1}{2}({}_{g}x_{.h} + {}_{h}x_{.g}), \tag{30}$$

which rearranged becomes

$$d_{..} - d_{gh} = \tfrac{1}{2}({}_{g}x_{.h} + .x_{.g} + {}_{h}x_{.g} + .x_{.h}). \tag{31}$$

When $g = j$, $h = k$, the comparative distance $h_{ik} = d_{gh} - d_{..}$. Since the x-values are functions of the observed proportions (Equation (6)), Equation (31) gives the comparative distances as functions of the observed data. Equation (31), then, gives a rather straightforward method for obtaining the best estimate, in a least-squares sense, of the matrix of comparative distances.

Obtaining Projections of Stimuli on Axes from Fallible Absolute Distances

For a situation in which the data are not fallible and in which absolute distances are given, Young and Householder have shown (a) how to determine if the stimuli lie in a real Euclidean space, (b) if they do, how to determine the dimensionality of the set of points, and (c) how to obtain the projections of the points on an arbitrary orthogonal reference system. This reference system may then be rotated to the "most meaningful" dimensions, if criteria for such are available.

We saw that if matrix B_i (Equation 5) is positive semidefinite, the stimuli lie in real Euclidean space. The rank of B_i (or two less than the rank of matrix F) is then equal to the dimensionality of the set of stimuli. Matrix B_i can be factored to obtain projections of the stimuli on an arbitrary set of orthogonal axes.

Matrix B_i, however, is constructed by placing the origin arbitrarily at one of the stimuli. With errorless data, the results will be identical (except for the orientation of axes and location of the origin) for each of the n possible matrices $B_i (i = 1, 2, \ldots, n)$. With fallible data, however, each point is somewhat in error. Assuming a true rank considerably less than the number of points, each matrix B_i will yield different results. We would then have the problem of deciding which B_i matrix gives the best solution.

One solution to this problem would be to place the origin at the centroid of the stimuli. This procedure would give a unique solution and would tend to allow the errors in the individual points to cancel each other. An origin at the centroid would, on the average, be less in error than an origin at any arbitrary stimulus. The problem would seem to be to find a convenient method of obtaining a matrix B^* with origin at the centroid of the stimuli instead of at one of the stimulus points.

We shall use the following notation:

m = axes $(m = 1, 2, \ldots, r)$,
j, k = points $(j, k = 1, 2, \ldots, n)$,
i = point taken as the origin,
a_{jm} = projection of point j on axis m, and
d_{jk} = distance between points j and k;

and take as given Equations (1) and (5):

$B_i = AA'$
$b_{jk} = \frac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2)$, where point i is taken as the origin. From Equation (5) it is seen that

$$b_{jk} = \sum_m^r a_{jm}a_{km} . \tag{32}$$

We shall, however, consider B_i to be an $n \times n$ matrix with the ith row and column composed of zero elements. In like manner A is $n \times r$ with the ith row composed of zero elements.

We wish to translate the axes from an origin at point i to an origin at the centroid of all n points.

Let $A^* = \| a_{jm^*} \|$ be the desired matrix of projections of points j on axis m^* of the new coordinate system with origin at the centroid of the n points.
Then

$$a_{jm^*} = a_{jm} - c_m , \tag{33}$$

where

$$c_m = \frac{1}{n}\sum_{j=1}^n a_{jm} = \text{the average projection of points on} \tag{34}$$

axis m = projection of centroid on axis m.

$$B^* = A^*A^{*\prime} = \| b_{jk}^* \|, \tag{35}$$

and

$$b_{jk}^* = \sum_m^r a_{jm^*}a_{m^*k} . \tag{36}$$

Substituting, we have

$$b_{jk}^* = \sum_m^r (a_{jm} - c_m)(a_{km} - c_m)$$

$$= \sum_m^r a_{jm}a_{km} - \sum_m^r a_{jm}c_m - \sum_m^r a_{km}c_m + \sum_m^r c_m c_m \,. \tag{37}$$

From Equation (34) it is seen that

$$b_{jk}^* = \sum_m^r a_{jm}a_{km} - \frac{1}{n}\sum_m^r a_{jm}\sum_i^n a_{im} - \frac{1}{n}\sum_m^r a_{km}\sum_k^n a_{km}$$

$$+ \frac{1}{n^2}\sum_m^r \Big[\sum_i^n a_{im}\Big]\Big[\sum_k^n a_{km}\Big]. \tag{38}$$

But

$$\sum_i^n b_{ik} = \sum_m^r a_{km}\sum_i^n a_{im} \tag{39}$$

and

$$\sum_k^n \sum_i^n b_{ik} = \sum_m^r \Big[\sum_i^n a_{im}\Big]^2. \tag{40}$$

Substituting, we have

$$b_{jk}^* = b_{ik} - \frac{1}{n}\sum_i^n b_{ik} - \frac{1}{n}\sum_k^n b_{ik} + \frac{1}{n^2}\sum_i\sum_k b_{ik} \,. \tag{41}$$

Equation (41) gives a routine method of translating a matrix B_i with origin at point i to an equivalent matrix B^* with an origin at the centroid of the points. It makes no difference, of course, which of the n matrices B_i is used in obtaining matrix B^*.

Matrix B^*, then, is the B-matrix we wish to factor to obtain projections of stimuli on axes.

Estimating $d_{..}$, the Unknown Additive Constant

The procedures of obtaining dimensionality and projections on axes discussed in the preceding section require absolute distances as given data. When the given data are comparative distances ($\hat{h}_{ik} = d_{..} - d_{ik}$)†, the unknown constant must be estimated to convert the comparative distances into absolute distances. We shall first consider the case where the data are not fallible, after which we shall discuss procedures for fallible data.

1. *Estimating $d_{..}$ from errorless comparative distances*

With errorless data, in order that the stimuli be considered as lying in a real Euclidean space of r dimensions, the B_i matrix must be positive semi-

†Comparative distances with signs reversed, actually ($h_{jk} = -\hat{h}_{jk}$).

definite and have a rank equal to r. This is equivalent to the statement that r latent roots of B_i must be positive and the remaining $(n - r)$ equal to zero.

The value of $d_{..}$ desired is the value which will permit the location of the stimuli in a real, Euclidean space of the smallest possible dimensionality. In terms of the matrix B^*, the value of $d_{..}$ desired is that value which results in the positive semidefinite B^* with the lowest rank. In terms of the latent roots, this becomes that value which results in a matrix B^* with the largest number of zero roots under the condition that the remaining non-zero roots are all positive.

This value can be determined, although it involves a tremendous amount of labor. The straightforward solution would be as follows:

1. Construct matrix B^* from the given data ($d_{ik} = d_{..} - \hat{h}_{ik}$).
2. Obtain the characteristic equation:

$$| B^* - \lambda I | = 0.$$

3. Set the last term equal to zero and solve for the real, positive values of $d_{..}$. This term will be an $(n - 1)$th degree polynomial in $d_{..}$. One of these values is the value desired.
4. Substitute each of the values for $d_{..}$ in the complete characteristic equation. Inspection of these equations shows which value of $d_{..}$ yields the largest number of zero roots.
5. The value which yields the largest number of zero roots with the remaining roots all positive is the value desired.

A "short-cut" procedure would be to evaluate the determinant of B^* directly. This determinant is the last term of the characteristic equation. One could then obtain the real, positive values of $d_{..}$ as in (3) above. Each value could then be substituted for $d_{..}$ in B^*. The latent roots of B^* could then be computed for each real positive value of $d_{..}$. One would be the desired value. This method would also involve a prohibitive amount of labor.

A third method would be to first estimate the dimensionality of the set of stimuli. To check the estimate, one could obtain an estimate of $d_{..}$ by evaluating one (or more) of the principal minors of B^* having an order equal to one greater than the estimated dimensionality. This estimate could then be substituted into B^* and the latent roots calculated.

There are a number of other methods possible involving the principal minors of the B^* matrix.† In general, they would all hinge on the fact that the correct estimate of $d_{..}$, if used in B^* results in (a) all principal minors of

†One could also use the matrix F (Equation 4) to obtain the value of $d_{..}$ which would minimize the dimensionality of the set of stimuli since it is known that the rank of F is two greater than the rank of B_i. There would seem to be little point in this, however, since F is a larger matrix, therefore involving more tedious procedures in evaluating the determinants, and since no properties of F have been given from which to determine whether the value of $d_{..}$ obtained will allow the stimuli to be placed in a *real* space.

order greater than the dimensionality of the stimuli being null, and (b) all principal minors of order equal to or less than the dimensionality being non-negative.

It is interesting to note that unless all points are equidistant from each other (in which case things collapse down to zero dimensions) it is always possible to obtain an estimate of $d_{..}$ in which the rank of B^* is at least two less than the number of stimuli. Thus before one could place any confidence in the obtained dimensionality of the stimuli, the rank of B^* would have to be smaller by three than the number of stimuli and preferably considerably smaller.

All of these methods require a great deal of labor. In addition, when fallible data are used, it is doubted whether the methods would give the solution we wish. *We would probably never obtain a positive semidefinite B^* matrix with a rank less than the number of stimuli minus two from fallible data.*

2. *Estimating $d_{..}$ from fallible comparative distances*

If we could obtain a positive semidefinite B^* whose non-zero roots consisted of a few large positive roots and a number of small positive roots by the methods outlined in the previous section, we could probably discard the small roots and conclude that the true dimensionality is equal to the number of large positive roots. Even in this case, however, there would probably be a better estimate of $d_{..}$. In the above example we have essentially assumed that any error must be such as to change the zero roots to positive values. It would be more reasonable to assume that errors would tend to change some zero roots in the positive direction and some in the negative direction. If we think of 3 points lying in a line so that $d_{12} + d_{23} = d_{13}$, the former would hold that any error would tend to make $(d_{12} + e_{12}) + (d_{23} + e_{23}) > (d_{13} + e_{13})$, whereas the latter would hold that $(d_{12} + e_{12}) + (d_{23} + e_{23}) < (d_{13} + e_{13})$ is equally likely.

This means that with fallible data the condition that B^* be positive semidefinite as a criterion for the points' existence in real space is not to be taken too seriously. What we would like to obtain is a B^*-matrix whose latent roots consist of

1. A few large positive values (the "true" dimensions of the system), and
2. The remaining values small and distributed about zero (the "error" dimensions).

It may be that for fallible data we are asking the wrong question. Consider the question, "For what value of $d_{..}$ will the points be most nearly (in a least-squares sense) in a space of a given dimensionality?" When one is interested in, or has reason to suspect, a one-dimensional case, the best $d_{..}$ in a least-squares sense is rather easy to obtain. In a one-dimensional set of points, $d_{ij} + d_{jk} = d_{ik} + e$ where j is between k and i, and e is an error.

In terms of available data ($d_{..} - d_{jk} = \hat{h}_{jk}$), this becomes

$$d_{..} - \hat{h}_{ij} + d_{..} - \hat{h}_{jk} = d_{..} - \hat{h}_{ik} + e$$

or

$$d_{..} + \hat{h}_{ik} - \hat{h}_{ij} - \hat{h}_{jk} = e.$$

The $d_{..}$ which will minimize the sums of squares of all of the e's would seem to be the $d_{..}$ desired. If we let

$$2F = \sum_{k>j>i} (d_{..} + \hat{h}_{ik} - \hat{h}_{jk} - \hat{h}_{ij})^2, \tag{42}$$

then, to minimize $2F$, we take the derivative of F with respect to $d_{..}$ and set it equal to zero. Designating this derivative as F', we have

$$F' = \sum_{k>j>i} (\hat{h}_{ik} - \hat{h}_{jk} - \hat{h}_{ij}) + \sum_{k>j>i} d_{..} = 0, \tag{43}$$

which rearranged becomes

$$\sum_{k>j>i} d_{..} = \sum_{k>j>i} (\hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik}). \tag{44}$$

Dividing by $\dfrac{6}{n(n-1)(n-2)}$, we find

$$d_{..} = \frac{6}{n(n-1)(n-2)} \sum_{k>j>i} (\hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik}). \tag{45}$$

In the one-dimensional case, it will ordinarily be possible to obtain the order of the n stimuli. If we define a symmetric matrix H_{jk} (j designates rows, k designates columns, $j, k = 1, 2, \cdots, n$) composed of elements $\hat{h}_{jk}$, the sum of the columns of H_{jk} divided by $(n - 1)$ gives the average distance of all points from each other minus the average distance from point k to all other points. Small values of $\hat{h}_{.k}$ indicate that k is near one end of the continuum, and large values indicate that k is near the center. Inspection of H_{jk} will ordinarily suffice to determine on which side of the continuum the particular stimulus is located. Given matrix H_{jk} with rows and columns in correct order, a shortcut method of obtaining

$$\sum_{k>j>i} (\hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik}) = L$$

is to

1. Obtain the diagonal sums S_c of elements above the principal diagonal:

$$S_c = \sum_{i=1}^{n-c} \hat{h}_{i(i+c)} \qquad (c = 1, 2, \cdots, n-1). \tag{46}$$

2. Multiply S_c by $(n - 2c)$. The sum is equal to L—i.e., if we let $t_c = (n - 2c)$,

$$L = \sum_{c=1}^{n-1} S_c t_c \,. \tag{47}$$

For the case where the dimensionality is expected to be greater than one, this general approach does not seem to be very practical. While one could think of finding that $d_{..}$ which will minimize the sums of squares of volumes of all possible tetrahedrons for the two-dimensional case, and the corresponding hyper-volumes for the higher-dimensional cases, it would seem that the labor involved would again be prohibitive.

There is another procedure which might serve to give a fair estimate of $d_{..}$ for cases where the expected dimensionality is greater than one. If a one-dimensional subspace of four or more points exists in the data, that subspace could be used to estimate $d_{..}$. While this procedure does not give a "best fit" in the least-squares sense, it does appear to be the most practical method suggested thus far. The method has been applied to actual data and was found to work quite well. The existence of such a subspace is relatively easy to determine. One can compute, for each set of three stimuli, the value of $d_{..}$ which would be required to locate the set of three along a straight line. There are $n(n-1)(n-2)/6$ of these "estimates," one for each set of three different stimuli; and they will be designated as $\tilde{d}_{..}$. The values of $\tilde{d}_{..}$ may be obtained from the following equation:

$$\tilde{d}_{..} = \hat{h}_{ii} + \hat{h}_{jk} - \hat{h}_{ik}\,, \qquad \text{where} \qquad \hat{h}_{ik} < \hat{h}_{ii}\,, \hat{h}_{jk}\,. \tag{48}$$

Given the $n(n-1)(n-2)/6$ values of $\tilde{d}_{..}$, the following points can be noted:

a. Except for error, points most nearly in a straight line will give the largest value of $\tilde{d}_{..}$.

b. If the four sets of three of any four points give about the same "highest" value of $\tilde{d}_{..}$ in a consistent manner, we can conclude that the four points are in a one-dimensional subspace. This value of $\tilde{d}_{..}$ would then be the estimate of $d_{..}$ wanted.

c. If such a set is not found, the largest value of $\tilde{d}_{..}$ might still be worth trying as an estimate of $d_{..}$. Using this value is equivalent to assuming that of the set of points at least one group of three is approximately linear. If one constructs a B-matrix with one of the points at the origin using this estimate and then finds that the third-order principal minors involving these three points vanish (approximately) this value of $\tilde{d}_{..}$ is probably a good estimate. The entire B-matrix need not, of course, be constructed. One would need to evaluate only $(n-3)$ third-order principal minors involving only 3 $(n-2)$ distinct elements instead of the $(n-1)(n-2)/2$ elements in the complete matrix.

Summary

A set of methods for multidimensional scaling based on Richardson's original model (**3**) have been developed, including a least-squares solution for obtaining comparative distances, and routine procedures for estimating the

additive constant necessary to convert comparative distances to absolute distances and for obtaining projections of stimuli on axes when fallible absolute distances are given. An outline of the procedures developed is given below.

A Routine Procedure for Multidimensional Scaling

A. To obtain comparative distances by the complete method of triads.

1. Construct the n matrices ${}_kP_{ij}$ from the raw data.
2. Construct the corresponding matrices ${}_kX_{ij}$.
3. Obtain a row vector of averages of columns for each of the n matrices ${}_kX_{ij}$.

$$ {}_kx_{.j} = \frac{1}{n-1} \sum_i {}_kx_{ij} . $$

4. Construct matrix ${}_kX_{.j}$ composed of these row vectors (k designates row, j designates columns).
5. Obtain a row vector of averages of columns of ${}_kX_{.j}$.

$$ {}_.x_{.j} = \frac{1}{n} \sum_k {}_kx_{.j} . $$

6, Add the gth element of ${}_.X_{.j}$ to each element in the gth row of ${}_kX_{.j}$. Call this new Matrix G_{kj}. Matrix G_{kj} thus contains the elements $({}_gx_{.j} + {}_.x_{.g})$.

7. Average the symmetric elements of G_{kj} to obtain the symmetric matrix H_{jk}. Matrix H_{jk} is composed of the elements $\hat{h}_{jk} = d_{..} - d_{jk}$, the comparative distances (with a negative sign) between stimuli.

$$ \hat{h}_{gh} = \hat{h}_{hg} = \tfrac{1}{2}(g_{gh} + g_{hg}). $$

B. To obtain an estimate of $d_{..}$.

1. If the hypothesis of unidimensionality of stimuli seems reasonable:
   a. Arrange rows and columns of H_{jk} in order of magnitude of the stimuli by
      (1) Noting magnitudes of sums of columns of H_{jk}, and
      (2) Examining elements of H_{jk}.
   b. Obtain diagonal sums of H_{jk} above principal diagonal.

$$ S_c = \sum_j^{n-c} \hat{h}_{j(j+c)} . $$

   c. Multiply each S_c by $(n - 2_c)$ and sum the products to obtain L.

$$ L = \sum_{c=1}^{n-1} S_c \, (n - 2c). $$

d. Divide L by $n\ (n-1)\ (n-2)/6$ to obtain $d_{..}$.

$$d_{..} = \frac{L}{n(n-1)(n-2)/6}.$$

2. If it is reasonable to assume dimensionality greater than one with at least one set of four stimuli in a one-dimensional subspace:
   a. Obtain the $n(n-1)\ (n-2)/6$ values of $\tilde{d}_{..}$ assuming in turn that each set of three stimuli lie in a line.

$$\tilde{d}_{..} = \hat{h}_{ij} + \hat{h}_{jk} - \hat{h}_{ik},$$

$$\hat{h}_{ik} < \hat{h}_{ij}, \hat{h}_{jk}.$$

   b. The four sets of three of any four points lying in a line will give the same "highest" value of $\tilde{d}_{..}$ (except for error) in a consistent manner. If such a set is found, this value of $\tilde{d}_{..}$ will be a good estimate of $d_{..}$.
   c. If no such set is found, use the highest value of $\tilde{d}_{..}$ obtained as an estimate of $d_{..}$. Compute the necessary elements of a matrix B_i with one of the three points as the origin. Evaluate the $(n-3)$ third-order principal minors of B_i. If these minors all vanish (approximately) this $d_{..}$ is probably a good estimate.

C. To obtain projections of stimuli on axes.

1. Construct D_{jk}.

$$d_{jk} = d_{..} - \hat{h}_{jk}.$$

2. Construct B_i with origin at any stimulus i.

$$b_{jk} = \tfrac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2).$$

3. Obtain from B_i averages of
   a. Columns,

$$b_{.k} = \frac{1}{n}\sum_j b_{jk},$$

   b. Rows,

$$b_{j.} = \frac{1}{n}\sum_j b_{jk},$$

   c. And all elements,

$$b_{..} = \frac{1}{n^2}\sum_k \sum_j b_{jk}.$$

4. Construct matrix B^* with origin at the centroid of stimuli.

$$b_{jk}^* = b_{jk} - b_{.k} - b_{j.} + b_{..}.$$

5. Factor B^*, obtaining A^*_{jm}, the matrix of projections of stimuli j on arbitrary axes m.
6. Rotate and translate matrix A_{jm} to a meaningful set of dimensions if criteria for such are available.

REFERENCES

1. Gulliksen, Harold. Paired comparisons and the logic of measurement. *Psychol. Rev.* 1946, **53**, 199–213.
2. Klingberg, F. L. Studies in measurement of the relations among sovereign states. *Psychometrika*, 1941, **6**, 335–352.
3. Richardson, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659. (Abstract).
4. Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389.
5. Young, G., and Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, **3**, 19–22.

Manuscript received 5/24/52

Revised manuscript received 7/14/52

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

THE AVERAGE SPEARMAN RANK CORRELATION COEFFICIENT*

SAMUEL B. LYERLY

WASHINGTON, D. C.

A method is derived for finding the average Spearman rank correlation coefficient of N sets of ranks with a single dependent or criterion ranking of n items without computing any of the individual coefficients. Procedures for calculating the exact distribution of ρ_{av} for small values of N and n are described for the null case. The first four moments about zero of this distribution are derived, and it is concluded that for samples as small as $N = 4$ and $n = 4$ the normal distribution can be used safely in testing the hypothesis $\rho_{av} = 0$.

The Spearman rank correlation coefficient ρ (sometimes called the "rank difference" coefficient after one method of calculating it) has been in use among psychologists for about half a century (**8**). Recent researches by Hotelling and Pabst (**2**), Kendall and his co-workers (**4**, **5**, **6**, **7**), and others has stimulated interest in rank correlation methods, largely because of their usefulness as non-parametric procedures—i.e., to provide tests of the null hypothesis in cases where the population distribution of either or both variables is unknown. Kendall's recent book (**7**) contains an excellent summary of rank correlation, including ρ and his own alternative coefficient τ, which in some respects is superior to ρ.

A short-cut method for computing the average of the intercorrelations of N ranked series each consisting of n items has been known for many years (**3**, 218). This method may be expressed in the following formula:

$$\text{Average inter-}\rho = 1 - \frac{N(4n + 2)}{(N - 1)(n - 1)} + \frac{12 \sum S^2}{N(N - 1)(n^3 - n)}, \qquad (1)$$

where

N = number of sets of rankings,
n = number of ranks in each set, and
S = sum of rank numbers for a given object or stimulus.

Kendall and Babington Smith discuss this formula and present the exact distribution of the mean intercorrelation in the null case for several small values of N and n (**6**; see also **7**, Chs. 6 and 7). Kendall's "coefficient of

*This problem first came to the writer's attention in discussions with Dr. Dean J. Clyde.

concordance," W, is a simple function of the average rank intercorrelation so designed that W ranges from 0 to 1 as the degree of agreement among the sets of ranks ranges from no agreement at all to perfect agreement. Approximate tests of significance of W (and hence of the average intercorrelation) based upon the z (or F) distribution or upon χ^2 are suggested for use with larger values of N or n where the computation of the exact probabilities is excessively laborious.

Occasionally there may arise a problem in which we are not interested in the average *inter*correlation of N sets of ranks, but we are concerned with the average correlation of N sets of ranks with a single dependent or criterion ranking. For example, we may ask N individuals to rank independently n objects of art according to their merit, and we may have as a criterion variable the rank-order listing of one or more experts. It is the purpose of this paper to derive a short method of calculating the average of the N ρ's without computing each ρ individually, and to investigate the problem of the significance of the average ρ in the null case.

We shall let y_i ($i = 1, 2, \cdots, n$) be the criterion or dependent set of ranks and x_{ij} ($i = 1, 2, \cdots, n; j = 1, 2, \cdots, N$) be the rank number assigned to the ith stimulus by the jth individual. Then the square of the difference in ranks for a single judgment $(y_i - x_{ij})^2$ is equal to $y_i^2 - 2y_i x_{ij} + x_{ij}^2$. The sum of these squares over a given i is

$$\sum_{j=1}^{N} (y_i - x_{ij})^2 = Ny_i^2 - 2y_i \sum_{j=1}^{N} x_{ij} + \sum_{j=1}^{N} x_{ij}^2 , \tag{2}$$

and the corresponding sum of squares over the whole table is

$$\sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 = \sum_{i=1}^{n} \sum_{j=1}^{N} x_{ij}^2 - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right) + N \sum_{i=1}^{n} y_i^2 . \tag{3}$$

Here both $\sum \sum x^2$ and $N \sum y^2$ are N times the sum of squares of the first n natural numbers and hence equal to $Nn(n + 1)(2n + 1)/6$. Thus (3) reduces to

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 &= 2 \sum_{i=1}^{n} \sum_{j=1}^{N} x_{ij}^2 - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right) \\ &= \frac{2Nn(n + 1)(2n + 1)}{6} - 2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right), \end{aligned} \tag{4}$$

and the average sum of squares of rank differences over the N individuals is

$$\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N} (y_i - x_{ij})^2 = \frac{2n(n + 1)(2n + 1)}{6} - \frac{2 \sum_{i=1}^{n} \left(y_i \sum_{j=1}^{N} x_{ij} \right)}{N} . \tag{5}$$

Substituting this average value of the sum of squares of differences into the usual formula, $\rho = 1 - 6 \sum d^2/(n^3 - n)$, we have

$$\rho_{a\cdot} = 1 - \frac{6\left[2n(n+1)(2n+1)/6 - 2\sum_{i=1}^{n}\left(y_i \sum_{j=1}^{N} x_{ij}\right)/N\right]}{n^3 - n}$$

$$= 1 - \frac{2(2n+1)}{n-1} + \frac{12\sum_{i=1}^{n}\left(y_i \sum_{j=1}^{N} x_{ij}\right)}{N(n^3 - n)}. \qquad (6)$$

The following fictitious example will illustrate the computation of $\rho_{a\cdot}$:

y	x				Σx	$y\Sigma x$
	1	2	3	4		
1	3	1	2	2	8	8
2	1	2	3	1	7	14
3	4	4	4	3	15	45
4	2	3	1	4	10	40
			Sums:		40	107

Here $N = 4$ and $n = 4$, and

$$\rho_{a\cdot} = 1 - \frac{2(8+1)}{3} + \frac{12 \times 107}{4 \times 60} = .35.$$

It may easily be verified that $\rho_{a\cdot}$ is the product-moment correlation coefficient of all Nn pairs of criterion-judgment ranks. This is a consequence of the fact that the distributions of all subsamples (individuals' judgments) are identical and the criterion distribution is likewise identical for all individuals; hence the "total correlation" is equivalent to the mean correlation.

The distribution of $\rho_{a\cdot}$ in the null case for any values of N and n can be found exactly, although a considerable amount of arithmetic is involved if either N or n exceeds 7 or 8. Since $\rho_{a\cdot}$ is the average of n ρ's, we need only the distribution of ρ in the null case for the appropriate n, and then by ordinary combinatorial methods the distribution of means (or sums) of N samples from this "population" can be found. Perhaps the most difficult part of the process is finding the initial distribution of ρ itself, although Kendall (**5,7**) has described methods of finding it and has tabulated the exact distributions of ρ (or of $\sum d^2$) for $n = 3$ to $n = 8$, inclusive. We shall illustrate the method

of finding the distribution of ρ_{av} for $n = 3$ and $N = 2$ by starting with the distribution of ρ for $n = 3$ as given by Kendall.

The distribution of ρ in the null case for $n = 3$ is as follows:

ρ	Relative frequency
1.00	1
.50	2
.00	0
− .50	2
−1.00	1

It should be noted that this is a discrete distribution and that only the listed values of ρ are possible, since the sum of squares of differences of 3 pairs of integers can take only a limited number of values. It will also be noted that the distribution is symmetrical about zero. This will obviously be true of the average ρ in the null case.

To find the distribution of ρ_{av} for $N = 2$, we merely calculate the distribution of all possible combinations of 2 from the above table. A ρ_{av} of + 1.00 may be obtained in only one way, i.e., when both single ρ's are + 1.00. The next highest possible ρ_{av} is + .75, which is obtained when one ρ is + 1.00 and the other is + .50. The relative frequency of this combination is 4, since the sequence + 1.00, + .50 can occur in two ways and so can the sequence + .50, + 1.00. The complete distribution may be obtained readily by constructing a table as follows:

	1	2	0	2	1				
1	1	2	0	2	1				
2		2	4	0	4	2			
0			0	0	0	0	0		
2				2	4	0	4	2	
1					1	2	0	2	1
Sums:	1	4	4	4	10	4	4	4	1

In this table the entries are all the possible products of pairs of elements of the original distribution. Each row is displaced one space to the right of its predecessor so that the new distribution can be obtained by merely adding columns. The total frequency is $(n!)^N$; and, tabulating the distribution in terms of ρ_{av}, we have:

ρ_{av}	Relative frequency
1.00	1
.75	4
.50	4
.25	4
.00	10
− .25	4
− .50	4
− .75	4
−1.00	1
Total	36

For small values of n and N the above method for finding the distribution of ρ_{av} is quite feasible. For large n and N the method is cumbersome, largely because the initial distribution of ρ itself is difficult to obtain. In such circumstances it is natural to inquire whether a good approximation can be found which will be satisfactory.

It is known that for large n the distribution of ρ itself approaches normality, but it is not known precisely how large n should be in order to use the normal integral for testing purposes. Kendall (7), with some hesitation, suggests 20 as a minimum n for which the normal curve may safely be used, and proposes that $\rho\sqrt{(n-2)/(1-\rho^2)}$, treated as "Student's" t with $n-2$ degrees of freedom, provides a better test in the range $8 < n < 20$. (For $n \leq 8$ exact probabilities have been computed.)

In the present case, where we have N sets of ranks, it is reasonable to suppose that the approach to normality should be fairly rapid. Since ρ_{av} is the mean of N values of ρ, and since ρ is approximately normally distributed, it would follow from the Central Limit Theorem that, for fixed n, the distribution of ρ_{av} would approach the normal as N increases.

The variance of ρ in the null case is $1/(n-1)$. Its fourth moment about zero (2) is

$$\mu_4 = \frac{3(25n^3 - 38n^2 - 35n + 72)}{25n(n+1)(n-1)^3}.$$

Here $\beta_1 = 0$ (all odd moments are zero by virtue of the symmetry of the distribution), and

$$\beta_2 = 3 + \frac{24}{100}\left(\frac{36 - 5n - 19n^2}{n^3 - n}\right),$$

which approaches the normal value of 3 as n increases.

In samples of N from such a population, we calculate from well-known theorems:

$$\sigma^2 = \frac{1}{N(n-1)},$$

$$\mu_4 = \frac{3(25n^3 - 38n^2 - 35n + 72)}{25N^3n(n+1)(n-1)^3} + \frac{3(N-1)}{N^3(n-1)^2},$$

and hence

$$\beta_2 = 3 + \frac{24}{100N}\left(\frac{36 - 5n - 19n^2}{n^3 - n}\right).$$

Thus the distribution of ρ_{as} approaches normality, as judged from its first four moments, very rapidly as n and N increase. Table 1 lists β_2 for certain small values of n and N.

TABLE 1

β_2 for Small Values of n and N

n	N					
	1	2	3	4	5	6
2	1.00	2.00	2.33	2.50	2.60	2.67
3	1.50	2.25	2.50	2.62	2.70	2.75
4	1.85	2.42	2.62	2.71	2.77	2.81
5	2.07	2.54	2.69	2.77	2.81	2.85
6	2.23	2.61	2.74	2.81	2.84	2.87
7	2.35	2.67	2.78	2.84	2.87	2.89

From Table 1 it can readily be seen that the normal value of 3 for β_2 is approximated very nearly for even small values of N and n. As a check on the usefulness of the normal distribution for testing the null hypothesis, the exact distribution of ρ_{as} for $N = 4$ and $n = 4$ was calculated (these values are probably as low as most experimenters would ever need to use), and the .005, .010, .025, and .050 points determined both from the exact relative frequencies and by using the normal approximation based upon the variance $1/N(n-1)$. (The four significance points were chosen because they provide one-tail or two-tail tests of the null hypothesis at the 1% or 5% levels.) Table 2 lists the relative frequencies in the tail of this distribution. From Table 2 we see, incidentally, that the ρ_{as} of .35 obtained in the fictitious example above is not significant at the 5% level, since a value of .50 is required for the one-tail test and .60 for the two-tail test.

TABLE 2

A Portion of the Distribution of ρ_{av} for $N = 4$ and $n = 4$

ρ_{av}	Relative frequency	Cumulative frequency	% cumulative frequency
1.00	1	1	.000003
.95	12	13	.00004
.90	58	71	.0002
.85	160	231	.001
.80	347	578	.002
.75	704	1282	.004
.70	1194	2476	.007
.65	1852	4328	.013
.60	2885	7213	.022
.55	3968	11181	.034
.50	5544	16725	.050
.45	7196	23921	.072
.....			
Total Frequency = 331776			

Table 3 lists the significant values of ρ_{av} as calculated from the exact values in Table 2 and the significant values as estimated from the normal approximation.

TABLE 3

Significant Points of the Distribution of ρ_{av} for $N = 4$ and $n = 4$

Significance level	Normal Approximation		Exact value
	By formula	Next higher possible value	
.005	.744	.75	.75
.010	.671	.70	.70
.025	.566	.60	.60
.050	.475	.50	.50

From Table 3 we see that for N and n as low as 4, the use of the normal integral in testing the significance of ρ_{av} at the 1% or 5% levels, using either the one-tail or the two-tail test, results in the same decision with respect to accepting or rejecting the null hypothesis as the use of the exact distribution. Since ρ_{av} approaches normality even more closely with larger N and n, and since most experimental problems would involve larger N or n or both, we are on safe ground in concluding that the normal curve approximation is

appropriate in testing the null hypothesis of a zero correlation in the population (or, more explicitly, of an average ρ of zero with the criterion in the population of sets of ranks from which the sample of sets was drawn).

In the non-null case, i.e., when the population ρ_{av} is not zero, no exact significance test is known, since the distribution of ρ itself is unknown for such populations. Thus we cannot test the hypothesis that an observed ρ_{av} is a sample from a hypothetical population in which the average ρ is some value other than zero, nor can we make an exact test of the difference between two sample values of ρ_{av}. Such tests must await the development of feasible methods of calculating or approximating the distribution of ρ for non-zero population values.*

REFERENCES

1. Friedman, Milton. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
2. Hotelling, H., and Pabst, M. R. Rank correlation and tests of significance. *Ann. math. Statist.*, 1936, **7**, 29–43.
3. Kelley, T. L. Statistical method. New York: The Macmillan Co., 1923.
4. Kendall, M. G. The advanced theory of statistics, Vol. I, 4th Ed. London: Charles Griffin and Co., 1948.
5. Kendall, M. G., Kendall, S. F. H., and Babington Smith, B. The distribution of Spearman's coefficient of rank correlation. *Biometrika*, 1938, **30**, 251–273.
6. Kendall, M. G., and Babington Smith, B. The problem of m rankings. *Ann. math. Statist.*, 1939, **10**, 275–287.
7. Kendall, M. G. Rank correlation methods. London: Charles Griffin and Co., 1948.
8. Spearman, Charles. 'Footrule' for measuring correlation. *Brit. J. Psychol.*, 1906, **2**, 89–108.

Manuscript received 1/8/52

Revised manuscript received 6/13/52

*One of the pre-publication reviewers of this paper has pointed out that the normal deviate test of ρ_{av} is equivalent to a χ^2 test of a linear relationship among the sums of ranks. Similar to Friedman's χ_r^2 (1), which he developed as a test of differences among the sums of ranks (and hence as a test of "concordance"), a χ_r with one degree of freedom as a test of linearity reduces to $\rho_{av}/\sigma_{\rho_{av}}$.

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

THE ORTHOGONAL APPROXIMATION OF AN OBLIQUE STRUCTURE IN FACTOR ANALYSIS

BERT F. GREEN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY*

A procedure is derived for obtaining an orthogonal transformation which most nearly transforms one given matrix into another given matrix, according to some least-squares criterion of fit. From this procedure, three analytic methods are derived for obtaining an orthogonal factor matrix which closely approximates a given oblique factor matrix. The case is considered of approximating a specified subset of oblique vectors by orthogonal vectors.

Introduction

In factor analysis problems, it is common practice to rotate the factor matrix to positive manifold or simple structure by means of oblique transformations. Many of the standard rotation techniques use oblique rather than orthogonal rotations. On the other hand, although the use of correlated factors in factor analysis is currently widespread, there is not complete accord on the issue. Some workers prefer to use only orthogonal reference frames. In addition to personal preferences, there may be other situations in which it is desired to have the final rotated factors orthogonal. When the results of the analysis are to be used in further mathematical formulas, such as in estimating factor scores (6) (11, Ch. 21), or in deriving multiple regression weights (3) (7) (9), orthogonal factors represent a simplification.

Thus the problem has arisen† of finding a set of orthogonal reference vectors which closely approximate a given oblique structure. This paper presents three methods by which this orthogonal reference frame can be determined analytically. These methods are all special cases of the general problem of a "best-fitting" orthogonal transformation.

It is of course possible to obtain an orthogonal reference frame by rotating the oblique solution into an orthogonal frame, using standard graphical rotation procedures. This relies heavily on subjective judgment. One advantage of an analytic method over such a procedure is that indeterminacy is eliminated—a single answer is always obtained. Also, for more than three or four factors, rotating to an orthogonal structure may involve a

*Part of this research was carried out while the author was a psychometric fellow at the Educational Testing Service, Princeton, New Jersey.

†This problem was first brought to the attention of the author by Dr. Dorothy C. Adkins.

great deal of cutting and trying. On the other hand, if compact computational methods are used, the computation involved in the analytic methods is not overwhelming. It is the author's opinion that the analytic methods will probably turn out to be more efficient than the "cut and try" methods. In fact, one of the main purposes in considering the problem was to discover an efficient means of fitting an orthogonal transformation.

A simple procedure for obtaining an almost orthogonal transformation has recently been proposed by Gibson (**5**). He suggests that in many cases the normalized centroids of corresponding reference and primary axes will form a set of vectors that are almost mutually orthogonal. This is certainly a very useful device and will be sufficient for some problems. However, it will seldom yield exact orthogonality. If precision is required, additional graphical rotations will be necessary. Also, there are many configurations in which the proposed set of centroids will not approach orthogonality. In such cases more complex methods are necessary.

General Problem

Solution for the General Problem

Let A and B be $k \times m$ matrices, with $k \geq m$, such that $A'B$ is of rank m, and consider the problem of finding the $m \times m$ orthogonal transformation Λ ($\Lambda'\Lambda = I$) which transforms A into B. Since in general it will not be possible to find a Λ which satisfies the equation

$$A\Lambda = B, \tag{1}$$

it is necessary to find an orthogonal transformation Λ such that the equation is satisfied as nearly as possible; that is, so that $A\Lambda$ is as close as possible to B, in some sense.

One definition of "closest" is the minimum of the sum of squared deviations. In this sense, it is required to find the Λ which minimizes the sum of squares of elements in the difference matrix $(A\Lambda - B)$.

Mosier (**8**) considered a similar problem where Λ is not restricted to an orthogonal matrix, but may be any non-singular transformation. He also dealt with the problem where the column vectors composing the transformation matrix had unit length. In the problem considered in this paper, it is required that the vectors be of unit length and that the scalar product between any two different vectors be zero.

Let

$$A = || a_{iq} ||; \qquad A_q = q\text{th column of } A;$$
$$B = || b_{is} ||; \qquad B_s = s\text{th column of } B;$$
$$\Lambda = || \lambda_{qs} ||; \qquad \Lambda_s = s\text{th column of } \Lambda;$$
$$\Theta = || \theta_{qh} ||; \qquad \Theta_q = q\text{th row of } \Theta;$$

where $i = 1, 2, \cdots, k$; $g, s, h = 1, 2, \cdots, m$; $k \geq m$. It is specified that $A'B$ be of rank m. The conditional equation on Λ is

$$\Lambda\Lambda' = \Lambda'\Lambda = I. \tag{2}$$

Using the method of Lagrange to impose condition (2), we minimize the function

$$2f = \sum_{s=1}^{m} \sum_{i=1}^{k} \Big(\sum_{g=1}^{m} a_{ig}\lambda_{gs} - b_{is} \Big)^2 + \sum_{g,h=1}^{m} \theta_{gh} \Big(\sum_{p=1}^{m} \lambda_{gp}\lambda_{hp} - \delta_{gh} \Big). \tag{3}$$

Here $\theta_{gh} = \theta_{hg}$ is a Lagrangian multiplier; δ_{gh} is the Kronecker delta ($\delta_{gh} = 0$, for $g \neq h$; $\delta_{gg} = 1$); the factor of 2 is introduced for convenience only. The partial derivative of f with respect to λ_{qs} is

$$\frac{\partial f}{\partial \lambda_{qs}} = \sum_{i=1}^{k} \Big[\Big(\sum_{g=1}^{m} a_{ig}\lambda_{gs} - b_{is} \Big) a_{iq} \Big] + \sum_{h=1}^{m} \theta_{qh}\lambda_{hs} \,. \tag{4}$$

Equating (4) to zero and using matrix notation we have

$$A_q'(A\Lambda_s - B_s) + \Theta_q\Lambda_s = 0. \tag{5}$$

The complete set of equations is then

$$A'(A\Lambda - B) + \Theta\Lambda = 0. \tag{6}$$

This may be rewritten as

$$(A'A + \Theta)\Lambda = A'B. \tag{7}$$

The transpose of (7) is

$$\Lambda'(A'A + \Theta) = B'A. \tag{8}$$

Postmultiplying (7) by (8) and using the conditional equation (2), we get

$$(A'A + \Theta)^2 = A'BB'A. \tag{9}$$

Taking the square root of each side of (9), we have

$$(A'A + \Theta) = (A'BB'A)^{\frac{1}{2}}. \tag{10}$$

Substituting (10) into (7) and solving for Λ, we have

$$\Lambda = (A'BB'A)^{-\frac{1}{2}}A'B. \tag{11}$$

Consider $(A'BB'A)^{-\frac{1}{2}}$. The fractional exponent is defined as follows: $Y = X^{\frac{1}{2}}$ if and only if $Y^2 = X$; by $X^{-\frac{1}{2}}$ is meant Y^{-1}. In general Y is not unique. Now for any square symmetric matrix of rank equal to its order, such as $A'BB'A$, it is possible to find an orthogonal matrix P and a diagonal matrix D such that

$$A'BB'A = PDP'. \tag{12}$$

The columns of P are the latent vectors (principal axes) of $A'BB'A$, and the diagonal elements in D are the corresponding latent roots (**1, 73**). Since $A'B$ is of rank m and of order $m \times m$, $A'BB'A$ is of rank m, and all roots are non-zero. Furthermore, since the product of a real matrix by its transpose yields a Gramian matrix, all roots are positive. If the latent roots are distinct, then P is unique except for the order of its columns. In a real symmetric matrix, the occurrence of equal roots introduces some indeterminacy in the matrix P. It is however always possible to find a P which satisfies (12). It can be shown that all P's which satisfy (12) will yield identical Λ's.

From (12) we have simply

$$(A'BB'A)^{\frac{1}{2}} = PD^{\frac{1}{2}}P'. \tag{13}$$

To verify this, note that

$$PD^{\frac{1}{2}}P'PD^{\frac{1}{2}}P' = PDP'. \tag{14}$$

It follows that

$$(A'BB'A)^{-\frac{1}{2}} = PD^{-\frac{1}{2}}P'. \tag{15}$$

Since either the positive or negative square root may be taken for each diagonal element in $D^{-\frac{1}{2}}$, there are 2^m solutions represented by equation (15). It can be shown that the solution which maximizes (3) is obtained by taking all square roots as positive.

Thus Λ is obtained in the following way:

a. The latent roots and vectors of $A'BB'A$ are determined. A simple, compact computational procedure for doing this has recently been developed by Bryan (**2**). Other methods may be found in Dwyer (**4**).

b. These roots and vectors are formed into the matrices D and P respectively, so that the root in the ith diagonal cell of D has its corresponding vector in the ith column of P, i.e., so that equation (12) holds.

c. $D^{-\frac{1}{2}}$ is computed by taking the reciprocals of the positive square roots of the corresponding diagonal elements in D.

d. Equations (15) and (11) are used to determine Λ.

Weights

Arbitrary fixed positive weights w_i^2 may be introduced in the function to be maximized, equation (3). If we simply replace a_{ig} by $w_i a_{ig}^*$, and b_{is} by $w_i b_{is}^*$, the function would become

$$f^* = \sum_{s=1}^{m} \sum_{i=1}^{k} w_i^2 \Big(\sum_{g=1}^{m} a_{ig}^* \lambda_{gs} - b_{is}^* \Big)^2 + \sum_{g=1}^{m} \sum_{h=1}^{m} \theta_{gh} \Big(\sum_{p=1}^{m} \lambda_{gp} \lambda_{hp} - \delta_{gh} \Big). \tag{16}$$

Here we initially have the matrices A^* and B^*, and wish to find the orthogonal transformation which minimizes f^*, (16). Let us define a $k \times k$ diagonal

matrix of weights, W, whose ith diagonal entry is w_i. Then, in order to use the general solution, (11), we may simply define A and B by

$$A = WA^*, \qquad B = WB^*. \tag{17}$$

Equation (11) may then be used to find the required orthogonal transformation. It is perfectly acceptable to have some of the weights zero, as long as A and B, (17), satisfy the requirements specified in the general development; that is, $A'B$ must be of rank m.

Note that it is *not* possible to introduce weights w_s in the present solution. This would amount to weighting the various orthogonal vectors in Λ differentially. To do this would require a different solution from the one considered in this paper.

Application to Factor Analysis

Definitions for the Factor Analysis Case

In the special factor analysis problem, it is required to find the orthogonal structure closest in some sense to a given oblique structure. "Closest" may be defined in at least three different ways, each of which yields a solution which is a special case of the above general result.

Let the following matrices be defined:

F = original factor matrix, with k tests (rows), and m factors (columns).
V = final oblique factor matrix.
G = oblique transformation. The various columns of G give the direction cosines of the various reference vectors with respect to the original uncorrelated factors.
Λ = orthogonal transformation. $\Lambda'\Lambda = I$.
H = matrix of primary axes. The various rows of H give the direction cosines of the various primary axes with respect to the original uncorrelated factors.
D_p = matrix whose elements are cosines of angles between primary axes and oblique reference vectors and therefore a diagonal matrix.

The rth reference vector is the normal to the rth hyperplane; the rth primary axis is the intersection of the $(m - 1)$ hyperplanes excluding the rth. The following relationships exist:

$R = FF'$, where R is the correlation matrix with communalities in the diagonal,

$FG = V$,

and

$HG = D_p$.

Detailed discussions of oblique structure have been given by Tucker (10)

and Thurstone (11, Ch. 15), using slightly different notation. [Our G is Tucker's Λ and Thurstone's Λ; our H is Tucker's H and Thurstone's T; our D_p is Tucker's D and Thurstone's D. Gibson (5) uses Thurstone's notation.]

First Method

One criterion for selecting the orthogonal transformation is based on a consideration of the oblique factor matrix V and the approximating orthogonal factor matrix $F\Lambda$. We may require the orthogonal factor loadings of each test to be as nearly similar as possible to its oblique factor loadings, according to a criterion of minimum sum of squares. To do this, we would minimize the sum of squares of elements in the difference matrix $(F\Lambda - V)$. Letting F be A and V be B in (11), we have

$$\Lambda = (F'VV'F)^{-\frac{1}{2}}F'V; \tag{18}$$

$$\Lambda = (F'FGG'F'F)^{-\frac{1}{2}}F'FG. \tag{19}$$

There is an alternative geometrical interpretation of this criterion of fit. A transformation of F may be viewed either as a rotation of axes with respect to the fixed points (vectors), or as a transformation of points with respect to the original fixed axes. In the latter interpretation, one could require that the distances between the points defined by $F\Lambda$ and those defined by FG be as small as possible. For test j, let the point defined by $F\Lambda$ be e_j with coordinates e_{js} and let the corresponding point defined by V be v_j , with coordinates v_{js} . The square of the distance between points e_j and v_j is $\sum_{s=1}^{m}(e_{js} - v_{js})^2$. Minimizing the sum over j of these quantities is identical with minimizing the sum of squares of elements in the matrix $(F\Lambda - V)$.

This first method for obtaining Λ provides for least over-all change in the factor loadings, and is a natural criterion to consider. However, it is weighted by the particular tests in the battery, and would not be invariant for a different selection of tests. Also, it may be said in general that each test vector has a weight in rough proportion to its communality. Thus the tests with the most common factor variance play the largest role in determining the orthogonal structure.

This solution is also affected by the distribution of test vectors in the factor space. One way of viewing the criterion of fit used in this method is that for each test vector, the angle it makes with each new orthogonal axis is as nearly as possible equal to the angle it makes with the corresponding oblique reference axis. It is clear from this statement that nothing explicit is said about the angles between the reference axes and the orthogonal axes. These angles are determined by the distribution of test vectors. In most cases the tendency will be for these angles to be small. However, if many tests are grouped together in some part of the configuration, peculiar results may be obtained. If most of the test vectors are grouped in the center of

the configuration with only a few vectors describing the corners of the structure, then there might be large changes in the loadings of the tests in the corner, in order to keep the center group relatively unaffected.

It may be noted that a criterion of fit very similar to this first method would consider the descriptions of the tests in terms of linear combinations of primary axes. These linear combinations are given by VD_p^{-1} . Thus, one might require the orthogonal factor loadings to differ as little as possible from these linear coefficients. To do this, V would be replaced by VD_p^{-1} in equation (18).

$$\Lambda = (F'VD_p^{-2}V'F)^{-\frac{1}{2}}F'VD_p^{-1}. \tag{20}$$

If all the diagonal entries in D_p are equal, equation (20) reduces to equation (18).

Second Method

The difficulties of the first method of determining Λ indicate that it would be desirable to have a method which did not depend on the particular distribution of test vectors in the configuration. One such method is obtained by requiring the new orthogonal axes to be as close as possible to the oblique reference axes defined by G.

In this case, "close" may be defined by maximizing the squares of the cosines of angles between corresponding axes and simultaneously minimizing the squares of cosines between non-corresponding axes. The intercosines are given by $G'\Lambda$. Thus the function to be minimized is the sum of squares of elements in $(G'\Lambda - I)$. From (11),

$$\Lambda = (GG')^{-1}G. \tag{21}$$

It should be noted that this is equivalent to maximizing the sum of the cosines of angles between corresponding axes, with no attention given to the angles between non-corresponding axes. In the preceding paragraph we minimized

$$f = \sum_{p=1}^{m}\sum_{q=1}^{m}\left(\sum_{s=1}^{m} g_{sq}\lambda_{sp} - \delta_{pq}\right)^2 + \theta, \tag{22}$$

where θ symbolizes the Lagrangian conditions, $G = \| g_{sq} \|$, $\Lambda = \| \lambda_{sp} \|$, and δ_{pq} is the Kronecker delta. Expanding (22) we get

$$f = \sum_{p}\sum_{q}\sum_{r}\sum_{s} g_{rq}g_{sq}\lambda_{rp}\lambda_{sp} - 2\sum_{p}\sum_{q}\delta_{pq}\sum_{s} g_{sq}\lambda_{sp} + m + \theta, \tag{23}$$

where all summations are from 1 to m. This reduces to

$$f = \sum_{q}\sum_{s} g_{sq}^2 - 2\sum_{q}\sum_{s} g_{sq}\lambda_{sq} + m + \theta. \tag{24}$$

Thus minimizing f with respect to λ_{ps} is the same as maximizing $\sum_q \sum_s g_{sq}\lambda_{sq}$, which is the sum of cosines between corresponding axes.

Third Method

Instead of requiring the orthogonal axes to be as close as possible to the reference axes, one can require them to be close to the primary axes. In this method we use the same definition of closeness that was used in the second method, with primary axes replacing reference axes. To obtain the solution under this criterion, H' is substituted for G in (21), which gives

$$\Lambda = (H'H)^{-\frac{1}{2}}H'. \tag{25}$$

Again it may be demonstrated that this result is equivalent to maximizing the sum of cosines between corresponding orthogonal and primary axes.

The author has not been able to find any simple analytic relationship between the second and third methods of solution. It is possible to show that if all diagonal entries in D_p are equal, that is, if all cosines of angles between corresponding *reference vectors* and *primary axes* are equal, then the solutions are identical. In this case, equation (25) reduces to equation (21). Since these cosines are usually quite similar, it seems likely that equations (21) and (25) will yield very similar results in most cases.

Weights

If for some reason it is desired to introduce differential weights, and if there is some analytic or subjective method for determining the weights, then the methods of section 3 may be used. In the first method of determining Λ in the factor analysis problem, we may weight the various tests differentially. Then, in the general solution, equation (11), $A = WF$, $B = WV$. In the second method, we may assign differential weights to particular reference vectors. In this case, $A = WG'$, $B = W$. In the third solution we may assign differential weights to different primary axes. In this case, $A = WH$; $B = W$.

The case of a subset of fixed vectors

So far we have been considering a problem with m factors for which an m-dimensional oblique structure has been determined. Methods have been developed for obtaining a set of m mutually orthogonal vectors which were closest to the oblique structure according to some least-squares criterion. In some applications it may be desired to fix some of the transformation vectors in advance, on the basis of some other criteria, and to determine only the remaining vectors by the methods of this paper.

In an oblique structure, some of the vectors may already be orthogonal among themselves and it may be desired not to change them. In other situations, there might be some oblique vectors which are so well determined by the configuration that it is decided to leave them unaltered and to use the orthogonal approximation only for the remaining vectors. In either case, r linearly independent vectors are fixed, thus determining a subspace of r dimensions, and it is required to find a set of orthogonal vectors which

span the complementary $(m - r)$ space. In either case, the methods of this paper could be used in the $(m - r)$ space orthogonal to the r fixed vectors. It is only necessary to assume that the fixed vectors correspond to particular vectors or factors in the structure, so that the $(m - r)$ orthogonal vectors to be determined match a particular set of $(m - r)$ factors or axes in the original structure. It is then possible to define the least-squares criterion in the reduced $(m - r)$ space, and to use the general results presented above.

Let us partition the transformation matrix Λ into the r fixed column vectors Λ_f, and the $(m - r)$ unknown column vectors Λ_u. Thus

$$\Lambda = || \Lambda_f \vdots \Lambda_u ||.$$

Here the restrictions are that $\Lambda_f' \Lambda_u = 0$, and $\Lambda_u' \Lambda_u = I$. Λ is not restricted to an orthogonal matrix since we do not need $\Lambda_f' \Lambda_f = 0$. The matrix V is also partitioned into the r columns V_f corresponding to Λ_f, and to the $(m - r)$ columns V_u corresponding to Λ_u. Likewise we have G_f and G_u, and H_f' and H_u'.

It is first necessary to project the structure onto the $(m - r)$ space which is orthogonal to the r fixed vectors. To do that we may use any set of $(m - r)$ orthogonal vectors denoted by the $m \times (m - r)$ matrix C, such that

$$C'C = I; \qquad \Lambda_f' C = 0. \tag{26}$$

These vectors merely define the $(m - r)$ subspace in which we will work.

It is a perfectly straightforward task to find a C which satisfies the restrictions specified by equations (26). However, this may be an arduous task if more than two or three vectors are fixed.

One method of obtaining C would be to determine each column successively. Calling the first column c_1, we let the last $(m - r - 1)$ elements of c_1 be zero. We then have $\Lambda_f' c_1 = 0$ as a set of homogeneous linear equations to solve for c_1. Then for c_2, we let the last $(m - r - 2)$ elements be zero, and solve the set of equations represented by $|| \Lambda_f \vdots c_1 ||' c_2 = 0$. In this method it is necessary to compute inverses of order $r \times r$,

$$(r + 1) \times (r + 1), \cdots, (m - 1) \times (m - 1).$$

Another method requires only inverses of order $(r \times r)$ or less. First an auxiliary matrix, U, of the same order as C is constructed. The matrix $|| \Lambda_f \vdots U ||$ is then broken up into submatrices according to the following schema:

$$\left\| \begin{array}{c|c|c|c} A & A^* & A & A \\ \hline B & X & B & B \\ \hline C & 0 & Y & C \\ \hline D & 0 & 0 & Z \end{array} \right\|$$

Here, $m = 3r + t$, A is of order $t \times r$; A^* is of order $t \times t$, made up of any t columns of A; X is of order $r \times t$; B, C, D, Y, and Z are of order $r \times r$. ($t \leq r$). We have

$$\Lambda_f = \left\| \begin{matrix} A \\ B \\ C \\ D \end{matrix} \right\|; \quad U_1 = \left\| \begin{matrix} A^* \\ X \\ 0 \\ 0 \end{matrix} \right\}; \quad U_2 = \left\| \begin{matrix} A \\ B \\ Y \\ 0 \end{matrix} \right\}; \quad U_3 = \left\| \begin{matrix} A \\ B \\ C \\ Z \end{matrix} \right\}.$$

Now orthogonality among blocks is obtained by writing the following equations:

$$\Lambda_f' U_1 = U_2' U_1 = U_3' U_1 = A'A^* + B'X = 0,$$

$$\Lambda_f' U_2 = U_3' U_2 = A'A + B'B + C'Y = 0,$$

and

$$\Lambda_f' U_3 = A'A + B'B + C'C + D'Z = 0.$$

Solving for the unknown submatrices we have

$$X = -(B')^{-1}A'A^*,$$

$$Y = -(C')^{-1}(A'A + B'B),$$

and

$$Z = -(D')^{-1}(A'A + B'B + C'C).$$

Next, for each U_i we must find a corresponding T_i such that $(U_iT_i)'\, U_iT_i = I$. To do this we form $U_i'\, U_i$ and factor completely, by the diagonal method, yielding

$$U_i'U_i = L_iL_i',$$

where L_i is triangular. Then $T_i = (L_i')^{-1}$. (The inverse of a triangular matrix may be determined very swiftly by solving in succession the equations represented by $TL' = I$.)

Then we have

$$C_1 = U_1T_1; \quad C_2 = U_2T_2; \quad C_3 = U_3T_3.$$

$$C = \| C_1 \mid C_2 \mid C_3 \|.$$

This method is general—it works for any m and r.

Now, in order to use the first criterion for determining Λ, (p. 434 above), we obtain FC. This gives the projections of the test vectors on the space defined by C. Using FC for A and V_u for B in the general solution (11), we obtain Λ_{FC}, an orthogonal transformation of order $(m - r) \times (m - r)$:

$$\Lambda_{FC} = (C'F'V_uV_u'FC)^{-\frac{1}{2}}C'FV_u. \tag{27}$$

Finally,

$$\Lambda_u = C\Lambda_{FC} . \tag{28}$$

To use the second method, (p. 435) we obtain $G_u'C$, which is used in place of G' in (21) to obtain Λ_{GC} .

$$\Lambda_{GC} = (C'G_uG_u'C)^{-\frac{1}{2}}C'G_u . \tag{29}$$

From this we have

$$\Lambda_u = C\Lambda_{GC} . \tag{30}$$

To use the third method of solution, (p. 436) we obtain H_uC which is used in place of H in (25) to obtain Λ_{HC} .

$$\Lambda_{HC} = (C'H_u'H_uC)^{-\frac{1}{2}}C'H_u' . \tag{31}$$

From this, we have

$$\Lambda_u = C\Lambda_{HC} . \tag{32}$$

A special case results when G_u (or H_u) is entirely within the $(m - r)$ subspace defined by C, that is, when the column vectors of G_u and those of C span the same space. In this case, since the subspace is defined by G_u , it is not necessary to determine C; (30) reduces to

$$\Lambda_u = G_u(G_u'G_u)^{-\frac{1}{2}}. \tag{33}$$

This case is equivalent to the case in which Λ_f is disregarded. (Either $\Lambda_f'G_u = 0$ and doesn't concern us, or Λ_f is simply disregarded.) Here the function maximized is the sum of cosines of angles between corresponding column vectors in Λ_u and G_u , subject to the restruction that $\Lambda_u'\Lambda_u = I$. Equation (33) can be derived directly from the maximization, or from (30).

When $\Lambda_u = \Lambda$, and $G_u = G$, it can be shown that (33) is equivalent to (21).

For H_u , we have in the special case,

$$\Lambda_u = H_u'(H_uH_u')^{-\frac{1}{2}}. \tag{34}$$

Summary

The problem has arisen in factor analysis of finding an orthogonal structure which approximates a given oblique structure. In order to solve this problem, a more general problem is considered; this is the problem of finding the orthogonal transformation which most nearly transforms one matrix into another, according to a least-squares criterion of fit. This general solution is represented by equation (11), with equation (15) defining one of the factors in (11). Arbitrary weights may be introduced in this general solution. For the factor analysis problem three analytic methods of determining the required orthogonal transformation are considered. The first method minimizes the sum of squared differences between oblique and orthogonal factor load-

ings. This solution depends on the size and distribution of test vectors in the configuration. The second method maximizes sum of cosines between corresponding reference and orthogonal axes. The third method maximizes the sum of cosines between corresponding primary and orthogonal axes. Arbitrary weights may be introduced in each method. This case is considered in which some of the axes are fixed in advance and the others are to be determined by the analytic methods developed in the paper.

Which of the methods is to be used in any problem depends on which criterion of closeness of fit is chosen by the investigator as being most appropriate for the particular problem.

REFERENCES

1. Aitken, A. C. Determinants and matrices. Edinburgh: Oliver and Boyd, 1942.
2. Bryan, J. G. A method for the exact determination of the characteristic equation and latent vectors of a matrix with applications to the discriminant function for more than two groups. Unpublished Ed. D. Thesis, Graduate School of Education, Harvard University, 1950.
3. Dwyer, P. S. The contribution of an orthogonal multiple factor solution to multiple correlation. *Psychometrika*, 1939, **4**, 163–171.
4. Dwyer, P. S. Linear computations. New York: John Wiley and Sons, 1951.
5. Gibson, W. Orthogonal and oblique simple structures. *Psychometrika*, 1952, **17**, 317–323.
6. Guilford, J. P., and Michael, W. B. Approaches to univocal factor scores. *Psychometrika*, 1948, **13**, 1–22.
7. Guttman, L. Multiple rectilinear regression and the resolution into components. *Psychometrika*, 1940, **5**, 75–100.
8. Mosier, C. I. Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 1939, **4**, 149–162.
9. Roff, M. Some properties of the communality in multiple factor theory. *Psychometrika*, 1936, **1**, 1–6.
10. Tucker, L. R. The role of correlated factors in factor analysis. *Psychometrika*, 1940, **5**, 141–152.
11. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.

Manuscript received 3/17/52

Revised manuscript received 4/25/52

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

BOOK REVIEW

Applied Statistics. A Journal of the Royal Statistical Society. L. H. C. TIPPETT, editor, Vol. 1, No. 1, March 1952. Oliver and Boyd, Ltd. London: 98 Great Russell Street, W. C. Single number 10s; Annual subscription 25s.

Control charts make their most important contribution in the field of personal relationships; in Great Britain, however, qualitity control (but not necessarily the use of statistics in industry) suffered a setback after the war; if x_1 is the reciprocal of the infant mortality rate per 1000, x_2 a logarithmic measure of "hereditaments" per head, x_3 illegitimacy rate per 1000 "related live births," y the percentage of households where the weekly basic wage of the senior wage earner is at least £ 7 10s, in a given town, then

$$y = 0.53\, x_1 + 0.67\, x_2 + 0.21\, x_3\,,$$

and a multiple correlation coefficient of .90 is found; the proportion voting conservative is not a good predictor; but the proportion of registered electors liable to jury service correlates well (from .6 to .8 or more) with the proportions in the "upper income group" in an administrative district; a factor analytic study shows Communists and Fascists to be alike very "tough-minded" though at opposite poles of a conservatism-radicalism scale, whereas Liberals lie in the middle of the latter scale but are extremely "tender-minded."

One should not infer, from the above list of oddments, that the new periodical *Applied Statistics* is intended merely as a compendium of miscellaneous information. Nevertheless, a heterogeneous assortment of topics is inevitable: The editor calls for papers of interest to "economists and social scientists, medical scientists and agricultural scientists, chemists and physicists, engineers and technologists." While psychologists are not mentioned as such, perhaps they are included as a species of the genus social scientist. In fact, this first issue does contain an interesting article by Hans J. Eysenck entitled "Uses and Abuses of Factor Analysis." At any rate, the diversity of interests to which Mr. Tippett hopes the journal will appeal is a reflection of the universality of statistical concepts.

Nevertheless, this universality is the universality of an abstraction, of mathematics, in fact, whereas *Applied Statistics* is to avoid mathematics. In the entire issue there are two pages devoted to a "mathematical derivation," besides three linear equations: the one given above, another in four, and one in two independent variables. The aim, as stated in a foreword by A. Bradford Hill, "is to present, in one way or another but always simply and clearly, the statistical approach and its value, and to illustrate in original articles modern statistical methods in their everyday applications." It is possible that in the endeavor to interest so many, the journal might prove to be of interest to none.

Beyond a doubt, Mr. Tippett is aware of the difficulty, and the first issue, at least, seems to avoid it very successfully. Though the sample is scarcely random, one can perhaps draw an inference, whatever may be the level of confidence. In addition to the articles, the Foreword, and an editorial, there are "Questions and Answers" (the editor promises to have the questions "answered by people competent to deal with the subjects raised"), "Notes and Comments," a book review, and reports and abstracts of addresses from the Industrial Applications Section of the Royal Statistical Society and from the Study Section of the Royal Statistical Society.

It seems appropriate to close this review with mention of the last of the abstracts. Mr. R. J. E. Silvey discussed a "survey of the structure of the television public and the effects of television on leisure." In households possessing TV, "Records were obtained of the behavior during one day of each individual." "A control sample was obtained by interviewing persons in households without television sets in the immediate vicinity. The findings are due to be published in the April issue of the *BBC Quarterly*."

Oak Ridge National Laboratory *A. S. Householder*

PSYCHOMETRIC SOCIETY

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1952

Receipts

Dues:

Year	Members	Student Members	
1952	359	43	
1951	26	6	
1949	1	..	
	386	49	$2,077.10*

Miscellaneous:

Psychometric Corporation	$5.00	
Duplicate Postal Note	3.00	
Unidentified Income	3.00	
		11.00
Total Receipts		$2,088.00

Disbursements

Psychometric Corporation (90% of dues)		$1,869.39
Miscellaneous Disbursements:		
Stationery and Postage	$120.22	
Secretarial and Clerical	87.45	
Addressograph Service	40.78	
Mimeographing	14.10	
Psychometric Corp. (check misaddressed)	5.00	
Bank Charges	.25	
		267.80
Total Disbursements		$2,137.19

Balance

Bank Balance, July 1, 1951	$1,096.11
Receipts	2,088.10
	$3,184.21
Expenditures, 1951–1952	2,137.19
Bank Balance, June 30, 1952	$1,047.02**

*Including overpayment of 10 cents
**Note deficit for 1952 of $49.09

PSYCHOMETRIC CORPORATION

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1952

Receipts

Subscriptions:

Year	No.	Amt. Pd.		
1951	28	$ 280.00		
1952	325	3,250.00		
1953	5	50.00		
	358	$3,580.00		$3,580.00
Psychometric Society				1,869.39
Sale of Back Issues				1,057.39
Other Income:				
Royalty—Psychometric Monographs			$ 3.76	
Dues Payments (for Psychometric Society)			8.00	
Overpayments and Other Income			18.00	
				29.76
Total Receipts				$6,536.64

Disbursements

Printing *Psychometrika:*		
Vol. 16, No. 2	$ 912.15	
Vol. 16, No. 3	1,057.46	
Vol. 16, No. 4	1,041.59	
Vol. 17, No. 1	1,421.31	
		$4,432.51
Printing Back Issues		1,115.52
Operating Expenses:		
Stipend of Asst. Editor	$337.50	
Secretarial Service, Office of Editor	600.00	
Stationery and Postage	145.84	
Treasurer's Bond	25.00	
Refunds—overpayments and cancellations	15.50	
Psychometric Society—dues sent to Corporation	8.00	
Agency Discounts	326.90	
		7,006.77

DISBURSEMENTS (Cont.)		
Psychometric Monograph:		
Printing	$2,889.05	
Editorial Fee	300.00	
Shipping to Chicago University Press	46.95	
		$3,236.00
Total Disbursements		$10,242.77
BALANCE		
Bank Balance, July 1, 1951		$9,144.89
Receipts, 1951–1952		6,536.64
		$15,681.53
Expenditures, 1951–1952		10,242.77
Bank Balance, June 30, 1952		$ 5,438.76

PSYCHOMETRIKA—VOL. 17, NO. 4
DECEMBER, 1952

MEMBERS OF THE PSYCHOMETRIC SOCIETY*

Sidney Adams, 242 North Granada Street, Arlington, Virginia
C. J. Adcock, Victoria University College, Wellington, New Zealand
Dorothy C. Adkins, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina
Roger B. Allison, Jr., Educational Testing Service, Princeton, New Jersey
Robert B. Ammons, Department of Psychology, University of Louisville, Louisville, Kentucky
Wade S. Amstutz, R.F.D. No. 1, Yellow Springs, Ohio
Anne Anastasi, Graduate School, Fordham University, New York 58, New York
O. F. Anderhalter, Bureau of Institutional Research, 15 N. Grand Boulevard, Saint Louis 3, Missouri
Scarvia B. Anderson, Apt. 606, 237 Audrey Lane, Washington 20, D. C.
T. Andrews, Department of Psychology, University of Maryland, College Park, Maryland
William H. Angoff, Educational Testing Service, Princeton, New Jersey
Valentine Appel, 3135 Decatur Avenue, New York 67, New York
Philip Ash, Industrial Relations Research, Inland Steel Co., 3210 Watling Street, East Chicago, Illinois
Melany E. Baehr, 137 S. Gilbert Street, Danville, Illinois
Andrew R. Baggaley, 5140 Woodlawn Avenue, Chicago 15, Illinois
Donald E. Baier, 5010 Brookeway Drive, Washington 16, D. C.
A. S. Barr, University of Wisconsin, Madison 6, Wisconsin
Brent Baxter, Agencies Research Division, Prudential Insurance Co. Home Office, Newark, New Jersey
David C. Beardslee, 531 S. Division Street, Ann Arbor, Michigan
Katherine Pease Beardsley, Beardsley Lane, Huntington, New York
Harold P. Bechtoldt, 2025 G Street, Iowa City, Iowa
Gordon Beckstead, 616 N. Third Street, Phoenix, Arizona
B. J. Bedell, 7 Third Avenue, Cascade, Saint Anns, Port-of-Spain, Trinidad, British West Indies
Ralph C. Bedell, 202 Lenoshire House, 3140 Wisconsin Avenue, N. W., Washington 16, D. C.
A. W. Bendig, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
George K. Bennett, Psychological Corporation, 522 5th Avenue, New York 18, New York
Joseph F. Bennett, 2134 Natural Science Building, University of Michigan, Ann Arbor, Michigan
Corlin O. Beum, Jr., Dunlap & Associates, Incorporated, 10 E. 49th Street, New York 17, New York
James E. Birren, 5550 S. Dorchester Avenue, Chicago 37, Illinois
Ruth Bishop, 49 W. 73 Street, Apt. b3, New York 23, New York
Reign H. Bittner, R.F.D. No. 1, Basking Ridge, New Jersey
Paul Blommers, State University of Iowa, Iowa City, Iowa
Frances Blumenthal, 6 Landscape Avenue, Yonkers 5, New York

*An asterisk preceding a name indicates student membership.

*Nicholas A. Bond, Jr., 3110 Greenfield Avenue, Los Angeles 34, California
Edward S. Bordin, Bureau of Psychological Service, 1027 East Huron Street, Ann Arbor, Michigan
Edgar F. Borgotta, 151 Central Street, Somerville, Massachusetts
Henry Borow, The General College, University of Minnesota, Minneapolis 14, Minnesota
Rev. W. A. Botzum, C.S.C., University of Portland, Portland 3, Oregon
*Eugene A. Bouvier, 3005 Hyperion, Los Angeles 27, California
*James E. Boyce, 919 N. Sixth Street, Lafayette, Indiana
Thomas L. Bransford, Examinations Division, State Department of Civil Service, Albany, New York
Robert Breden, Personnel Research Department, Sixth and Main Streets, Proctor and Gamble Co., Cincinnati, Ohio
Claude F. Bridges, 48 Post Avenue, New York 34, New York
Hubert E. Brogden, Personnel Research and Procedures Branch, AGO, Pentagon, Washington, D. C.
Cecil R. Brolyer, Division of Examinations, N. Y. State Department of Civil Service, 39 Columbia Street, Albany, New York
Daniel F. Brophy, Department of Student Life, College of the City of New York, 139th Street and Convent Avenue, New York 31, New York
Theodore H. Brown, Soldiers Field Road, Harvard Business School, Boston 63, Massachusetts
Mrs. Virginia M. Brown, 1700 East Tuft Avenue, Englewood, Colorado
Joseph G. Bryan, 74 Springfield Street, Watertown 72, Massachusetts
Edward P. Buckley, 1156 Nineteenth Street, N. W., Washington, D. C.
Hermann F. Buegel, Box 114, University of North Dakota, Grand Forks, North Dakota
Mrs. Laverne K. Burke, 1706B Commonwealth Avenue, Alexandria, Virginia
Paul J. Burke, 89-20 161 Street, Jamaica 2, New York
William W. Burke, School of Social Work, Washington University, Saint Louis, Missouri
Oscar K. Buros, 220 Montgomery Street, Highland Park, New Jersey
Harold E. Burtt, Department of Psychology, Ohio State University, Columbus 10, Ohio
Hobart Bushey, Hunter College, Lexington Avenue & 68th Street, New York 21, New York
Joel T. Campbell, Personnel Research and Procedures Branch, AGO, The Pentagon, Washington 25, D. C.
*Earl R. Carlson, 1465 University Terrace, Apt. 1315, Ann Arbor, Michigan
John B. Carroll, 40 Quincy Street, Cambridge 38, Massachusetts
Harold D. Carter, 10 E. Parnassus Ct., Berkeley 8, California
Launor F. Carter, Department of Psychology, University of Rochester, Rochester 3, New York
Teobaldo Casanova, Ave. Fernandez Juncos 1064, Santurce, Puerto Rico
Robert L. Chapman, Rand Corporation, 1500 Fourth Street, Santa Monica, California
Henry Chauncey, Educational Testing Service, Princeton, New Jersey
Stephen C. Clark, Department of Psychology, John Muir College, Pasadena 3, California
*Theodore W. Clymer, 717 14th Avenue, S. E., Minneapolis 14, Minnesota
Herbert H. Coburn, Box 92, University Station, Seattle 5, Washington
George H. Collier, Department of Psychology, Duke University, Durham, North Carolina
Andrew L. Comrey, Jr., Department of Psychology, University of California, Los Angeles 24, California
Herbert S. Conrad, Research and Statistical Service, U. S. Office of Education, Washington 25, D. C.
Lt. Comdr. Ellsworth B. Cook, 2 Pooks Hill Road, Bethesda, Maryland
C. H. Coombs, Department of Psychology, University of Michigan, Ann Arbor, Michigan
Herman A. Copeland, 543 Overlook Road, Philadelphia 28, Pennsylvania

Francis G. Cornell, College of Education, University of Illinois, Champaign, Illinois
Sarah Counts, 5737 Drexel Avenue, Chicago 37, Illinois
John T. Cowles, Educational Testing Service, Princeton, New Jersey
Gertrude M. Cox, North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina
H. M. Cox, University of Nebraska, Lincoln 8, Nebraska
Kenneth J. Cox, 159 Emerson Street, Hamilton, Ontario
Clarke W. Crannell, 120 S. Main Street, Oxford, Ohio
William J. E. Crissy, 51-01 39th Avenue, Sunnyside, Long Island City, New York, New York
Joan H. Criswell, 2728 32 Street, N. W., Washington 8, D. C.
Lee J. Cronbach, Bureau of Research and Service, College of Education, University of Illinois, Urbana, Illinois
Dorothea J. Crook, Psychology Department, Tufts College, Medford 55, Massachusetts
Elmer K. Culler, Laboratory of Psychology, River Campus, University of Rochester, Rochester 3, New York
E. E. Cureton, 1907 Lake Avenue, Knoxville 16, Tennessee
Louise W. Cureton, 1907 Lake Avenue, Knoxville 16, Tennessee
Manuel Cynamon, 10 Avenue P, Brooklyn 4, New York
W. Grant Dahlstrom, Department of Psychology, University of Iowa, Iowa City, Iowa
Karl M. Dallenbach, 2106 Meadowbrook Drive, Austin, Texas
Melvin D. Davidoff, 10615 Dunkirk Drive, Silver Spring, Maryland
Frederick B. Davis, Educational Clinic, Hunter College, 695 Park Avenue, New York 21, New York
*Paul C. Davis, c/o Roberts Wesleyan College, North Chili, New York
Hugh M. Davison, 419 Martin Terrace, State College, Pennsylvania
Walter L. Deemer, Jr., 3624 Appleton Street, N. W., Washington 8, D. C.
James W. Degan, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge 39, Massachusetts
Robert G. Demaree, Department of Psychology, University of Illinois, Urbana, Illinois
J. C. Denton, Jr., 2815 Digby Avenue, Cincinnati 20, Ohio
Fernando Rodrigues de Silveira, Rua Francisca Salles 90, Rio de Janeiro, Brazil
*Harvey F. Dingman, Psychology Department, University of Southern California, Los Angeles 7, California
Jerome E. Doppelt, 2165 Ryer Avenue, New York 57, New York
Paul L. Dressel, Michigan State College, East Lansing, Michigan
Philip H. DuBois, Department of Psychology, Washington University, Saint Louis 5, Missouri
Jack W. Dunlap, Lockwood Road, Riverside, Connecticut
Theodore F. Dunn, 4443 C Street, S. E., Washington, D. C.
Paul S. Dwyer, 403 Lenawee Drive, Ann Arbor, Michigan
Henry Dyer, 25 Dwight Street, Brookline 46, Massachusetts
Robert L. Ebel, Room 114 University Hall, Iowa City, Iowa
Harold Edgerton, 2 Nursery Road, New Canaan, Connecticut
Allen L. Edwards, Department of Psychology, University of Washington, Seattle 5, Washington
Daisy M. Edwards, 1 Oakfield Court, Queens Road, Weybridge, Surrey, England
Daniel R. Embody, Washington Water Power Co., P. O. Drawer 1445, Spokane 6, Washington
Wm. G. Emmett, Moray House, Edinburgh 8, Scotland
Max D. Engelhart, 5427 Greenwood Avenue, Chicago 15, Illinois
Sidney Epstein, P. O. Box 291, Ann Arbor, Michigan

Mrs. Katherine W. Estes, Department of Psychology, Indiana University, Bloomington, Indiana
Chester E. Evans, Employee Research Section, General Motors Corporation, Detroit 2, Michigan
Dorothea W. F. Ewers, 745 Aberdeen Drive, Lincolnshire Estates, Crete, Illinois
Harold T. Fagin, 88 Mitchell Avenue, Binghampton, New York
*Edmund W. Faison, 903 Dartmouth Drive, Bucknell Manor, Alexandria, Virginia
Mary E. Farabaugh, 4201 Massachusetts Ave., N. W., Washington 16, D. C.
Samuel W. Fernberger, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania
Warren G. Findley, Educational Testing Service, Princeton, New Jersey
John C. Flanagan, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pennsylvania
Ronald C. Force, 4310 S. Lipscomb, Amarillo, Texas
John W. French, Educational Testing Service, Princeton, New Jersey
Benjamin Fruchter, 1216 Norwood Road, Austin, Texas
Edmund F. Fuchs, 1806 Florin Street, Silver Spring, Maryland
Marguerite S. Gadel, Personnel Division, Prudential Insurance Company, Newark, New Jersey
N. L. Gage, Bureau of Research and Service, College of Education, University of Illinois, Urbana, Illinois
Henrietta L. Gallagher, 43 Wiggins Street, Princeton, New Jersey
William Thomas Gallagher, 32–82 42nd Street, Long Island City, New York
Eric F. Gardner, Syracuse University, Syracuse 10, New York
Henry E. Garrett, Department of Psychology, Columbia University, New York 27, New York
Frederick Gehlman, Shady Lane Flowerfield, Lombard, Illinois
Charles P. Gershenson, 255 Fir Street, Park Forest, Illinois
Jerome H. Gibson, 2957 Santa Clara Drive, Albuquerque, New Mexico
Wilfred A. Gibson, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina
Robert Glaser, American Institute for Research, 413 Morewood Avenue, Pittsburgh 13, Pennsylvania
Dr. Goldine C. Gleser, School of Medicine, Washington University, 640 South Kingshighway, Saint Louis 10, Missouri
Howard W. Goheen, 2803 16th Road South, Arlington, Virginia
*Maxim Goode, 6658 Whitley Terrace, Los Angeles 28, California
Leonard V. Gordon, Department of Psychology, University of New Mexico, Albuquerque, New Mexico
Mary Agnes Gordon, 128 Main Plaza, San Antonio, Texas
David A. Grant, Department of Psychology, University of Wisconsin, Madison 6, Wisconsin
Donald L. Grant, 76 Heatherbloom Road, White Plains, New York
J. Douglas Grant, 119 Gardner Street, Vallejo, California
Bert F. Green, Jr., 22–A–233c R.L.E., Massachusetts Institute of Technology, Cambridge 39, Massachusetts
Dorothy E. Green, Human Resources Research Laboratory, Bolling Air Force Base, Washington 25, D. C.
Russel F. Green, 119 University Park, Rochester 20, New York
*Richard B. Grekila, 2009 S. Vine Street, Muncie, Indiana
William W. Grings, Department of Psychology, University of Southern California, Los Angeles 7, California
*Byron Lou Groesbeck, 661 Swift Street, Apt. 4, Ann Arbor, Michigan

J. P. Guilford, P. O. Box 1134, Beverly Hills, California
Harold O. Gulliksen, Box 592, Princeton, New Jersey
John W. Gustad, University Counseling Service, Vanderbilt University, Nashville, Tennessee
Louis Guttman, Israel Institute of Applied Social Research, Shell Building, Jerusalem, Israel
Raef K. Haddad, 1052 Brown, New York University, Washington Square, New York 3, New York
Harold Wesley Hake, Institute for Cooperative Research, 1315 Saint Paul Street, Baltimore 2, Maryland
H. H. Harman, 4111 Maryland Drive, Washington 16, D. C.
Bertha P. Harper, Personnel Research and Procedures Branch, AGO, The Pentagon, Washington 25, D. C.
Charles M. Harsh, Human Factors Division, U. S. Navy Electronics Laboratory, San Diego 52, California
Edward N. Hay, 245 Ogden Avenue, Swarthmore, Pennsylvania
E. R. Henry, 428 Demarest Avenue, Oradell, New Jersey
Alfred Felix Hertzka, 1030 N. Sweetzer Avenue, Los Angeles 46, California
John Harry Hill, 925½ F Street, Coronado 18, California
M. A. Hiller, International Harvester Company, P. O. Box 179, Louisville, Kentucky
*John R. Hills, 4509 Halldale Avenue, Los Angeles 62, California
Elmer D. Hinckley, Box 2007, University Station, Gainesville, Florida
E. Lee Hoffman, Urban Life Research Institute, 127 Elk Place, Tulane University, New Orleans, Louisiana
Paul Horst, Department of Psychology, University of Washington, Seattle, Washington
Alston S. Householder, Oak Ridge National Laboratory, Oak Ridge, Tennessee
Cyril J. Hoyt, Bureau of Educational Research, University of Minnesota, Minneapolis, Minnesota
Doncaster G. Humm, Box 1433 Del Valle Station, Los Angeles 15, California
Lyman C. Hunt, Jr., School of Education, Pennsylvania State College, State College, Pennsylvania
Cuthbert C. Hurd, International Business Machines Corporation, 590 Madison Avenue, New York 22, New York
Mr. Tarow Indow, 1960 Isshiki, Hayama, Mieuragun, Kanagawaken, Japan
Mrs. Clifford L. Jackson, 246 Lexington Avenue, New York 16, New York
Rheem F. Jarratt, c/o Committee on Statistics, University of Chicago, Chicago 37, Illinois
Nathan Jaspen, National League for Nursing Education, 2 Park Avenue, New York 16, New York
*Rutledge Jay, Department of Personnel Methods, School of Business Administration, Wayne University, Detroit 1, Michigan
Charlotte Dragstedt Jeffrey, 9857 Hoxie Avenue, Chicago 17, Illinois
Thomas N. Jenkins, 105-21 66th Avenue, Forest Hills, New York
William Leroy Jenkins, Department of Psychology, Lehigh University, Bethlehem, Pennsylvania
William Alexander Jenner, 3426 78th Place, S. E., Washington 19, D. C.
Cecil Davis Johnson, 406 Fenwick Drive, Falls Church, Virginia
*Lt. Woodbury Johnson, MSC, USNR, U. S. N. School of Aviation Medicine, N. A. S. Pensacola, Florida
Palmer O. Johnson, University of Minnesota, Minneapolis 14, Minnesota
Lyle V. Jones, Department of Psychology, University of Chicago, Chicago 37, Illinois
Clifford E. Jurgensen, Minneapolis Gas Company, Minneapolis, Minnesota

Lessing A. Kahn, 12021 Milton Street, Wheaton, Maryland
Harry Kaplan, 1526 Fort Davis Street, S. E., Washington 20, D. C.
John E. Karlin, Bell Telephone Laboratories, Murray Hill, New Jersey
Lawrence Karlin, Psychology Department, WSC, New York University, Washington Square, New York, New York
Samuel Karvruck, 706 Buchanan Street, N. E., Washington 17, D. C.
Leo Katz, Department of Mathematics, Michigan State College, East Lansing, Michigan
Raymond Katzell, Psychological Service Center, Syracuse University, 125 College Place, Syracuse 10, New York
Frederick Kehew, P. O. Box 172, Atlantic City, New Jersey
Robert J. Keller, 211 Burton Hall, University of Minnesota, Minneapolis 14, Minnesota
T. L. Kelley, 1611 Paterna Road, Santa Barbara, California
*Harold E. Kerber, Chillicothe Road, Novelty, Ohio
Mary J. Kientzle, Department of Psychology, Washington State College, Pullman, Washington
Frank Killian, Jr., 3700 Medical Group, Lackland AF Base, San Antonio, Texas
Joseph E. King, Industrial Psychology, Inc., 105 W. Adams Street, Chicago 3, Illinois
*Walter A. Klieger, 5726 Fifth Street N. E., Washington 11, D. C.
William E. Kline, The Choate School, Wallingford, Connecticut
Helen Lois Koch, 1374 E. 57th Street, Chicago 37, Illinois
Leonard S. Kogan, Community Service Society of New York, 105 E. 22nd Street, New York 10, New York
Joseph F. Kubis, Fordham University, Graduate School, New York 58, New York
Frederic Kuder, College Station, Duke University, Durham, North Carolina
Albert K. Kurtz, Penn State Apts., No. 35, State College, Pennsylvania
George Kuznets, Giannini Foundation, University of California, Berkeley 4, California
Oliver L. Lacey, University of Alabama, University, Alabama
Thomas A. Lamke, Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa
Herbert D. Landahl, 5741 Drexel Avenue, Chicago 37, Illinois
Charles R. Langmuir, Slocum Hall 200, Syracuse University, Syracuse 10, New York
Wilbur L. Layton, Student Counseling Bureau, 101 Eddy Hall, University of Minnesota, Minneapolis, Minnesota
Paul F. Lazarsfeld, Department of Sociology, Columbia University, New York 27, New York
D. Welty Lefever, University of Southern California, University Park, Los Angeles 7, California
John M. Leiman, Personnel Research Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas
Vernon W. Lemmon, Washington University, St. Louis 5, Missouri
Bernard Levenson, 535 W. 111th Street, New York 25, New York
Hollis N. Leverett, 21 Hampden Street, Wallesley, Massachusetts
Ernest M. Ligon, Union College, Schenectady 8, New York
Fil Lic R. Lindahl, Ovra Husargaban 34, Goteborg, Sweden
E. F. Lindquist, College of Education, State University of Iowa, Iowa City, Iowa
Kenneth B. Little, Psychology Department, Stanford University, Stanford, California
Frederic Lord, 20 Pine Street, Princeton, New Jersey
Irving Lorge, Teachers College, Columbia University, New York 27, New York
Maurice Lorr, 1521 Erskine Street, Washington 12, D. C.
Constance Lovell, 663 W. 34th Street, Apt. 401, Los Angeles 7, California
Ardie Lubin, 1511 Franklin Street, N. E., Washington 18, D. C.
Bertha M. Luckey, Psychology Clinic, 1380 E. 6th Street, Cleveland 14, Ohio
George A. Lundberg, Department of Sociology, University of Washington, Seattle, Washington

Walter A. Lurie, National Community Relations Advisory Council, 9 E. 38th Street, New York 16, New York
Samuel B. Lyerly, 2908 30th Street, S. E., Washington 20, D. C.
Harold L. McAdoo, 2425 Central Avenue, Alexandria, Virginia
William A. McClelland, Detachment No. 2, Human Resources Research Center, Mather Air Force Base, California
*Robert L. McCormack, 112 Psychology, University of Minnesota, Minneapolis 14, Minnesota
Richard McHugh, Department of Psychology, Iowa State College, Ames, Iowa
Kenneth F. McLaughlin, 433 Dewey Drive, Annapolis, Maryland
Walter J. McNamara, International Business Machines Corp., 590 Madison Avenue, New York, New York
Quinn McNemar, Department of Psychology, Stanford University, California
Louis L. McQuitty, Department of Psychology, University of Illinois, Urbana, Illinois
Benjamin Malzberg, Department of Mental Hygiene, State Office Building, Albany 1, New York
*Eric Marder, 110 W. 84th Street, New York 24, New York
Melvin R. Marks, AGO Personnel Research Section, Department of the Army, Washington 25, D. C.
William G. Matheny, University of Illinois, Urbana, Illinois
Luzelle Denton Mays, 8311 Roanoke, Takoma Park, Maryland
Francis F. Medland, Personnel Department, Room 4030, Arlington Annex, Washington 25, D. C.
*Donald M. Medley, 4059 Lakeland, Robbinsdale 22, Minnesota
Paul E. Meehl, 112 Psychology Building, University of Minnesota, Minneapolis 14, Minnesota
Peter F. Merenda, 6570 Fifth Avenue, Kenosha, Wisconsin
Charles A. Metzner, School Public Health, University of Michigan, Ann Arbor, Michigan
Lorenz A. Meyer, 1814 Kanawha Street, Adelphi, Maryland
William B. Michael, Rand Corporation, 1500 4th Street, Santa Monica, California
William G. Mollenkopf, 112 Cedar Lane, Princeton, New Jersey
William T. Moran, Pillsbury Mills, Incorporated, Pillsbury Building, Minneapolis, Minnesota
Donald R. Morrison, 7681 B Washington Lane, Philadelphia 17, Pennsylvania
James N. Mosel, Department of Psychology, George Washington University, Washington 6, D. C.
Frederick Mosteller, 309 Emerson Hall, Harvard University, Cambridge 38, Massachusetts
Carl Murchison, The Journal Press, 2 Commercial Street, Provincetown, Massachusetts
*Barbara Jacobsen Myers, c/o Psychology Department, Yale University, New Haven, Connecticut
Paul M. Neurath, 510 W. 124th Street, New York 27, New York
Wilson L. Newman, 5539 S. Kenwood Avenue, Chicago 37, Illinois
*William T. Nicholas, 813 University Avenue S. E., Minneapolis 14, Minnesota
Raymond C. Norris, 20 East Hillwood Avenue, Orangeburg, New York
James A. Norton, Jr., Statistical Laboratory, Purdue University, Lafayette, Indiana
*Elbert F. Nothern, Blue Mountain College, Blue Mountain, Mississippi
Ewald B. Nyquist, R.F.D. No. 1, Ravena, New York
Ingram Olkin, Department of Mathematics, Michigan State College, East Lansing, Michigan
Willard C. Olson, University Elementary School, University of Michigan, Ann Arbor, Michigan
Beatrice Orleans, Apt. 522, 2032 Belmont Road, N. W., Washington, D. C.
Siroon Pashalian, 65 W. 181 Street, New York 53, New York

Richard Pearson, Educational Testing Service, Princeton, New Jersey
Edwin A. Peel, Department of Education, The University, Edmund Street, Birmingham 3, England
Donald C. Pelz, 909 E. Huron Street, Ann Arbor, Michigan
Robert Perloff, 3308 Buchanan Street, Apt. 103, Mount Ranier, Maryland
*Norman C. Perry, 333 Auzerais Street, San Jose, California
*Norman D. Peterson, Espanola, Washington
Shailer A. Peterson, American Dental Association, 222 E. Superior Avenue, Chicago 11, Illinois
Luigi Petrullo, 2431 N. Edgewood Street, Arlington, Virginia
B. Roger Philip, 323 Annette Street, Toronto, Canada
H. T. H. Piaggio, Department of Mathematics, University College, University Park, Nottingham, England
Henri Pieron, Institut de Psychologie a la Sorbonne, Rue des Ecoles, Paris 5, France
Joseph A. Pittman, North Carolina College, Durham, North Carolina
Lynette Birdsall Plumlee, Box 592, Princeton, New Jersey
James A. Rafferty, 817 Allen Street, Falls Church, Virginia
John H. Rapparlie, 9261 Grayton, Detroit 24, Michigan
Philburn Ratoosh, Department of Psychology, University of Wisconsin, Madison 6, Wisconsin
William Reitz, Wayne University, Detroit 1, Michigan
Martin L. Reymert, Mooseheart Laboratory for Child Research, Mooseheart, Illinois
Horacio J. A. Rimoldi, Universidad de la Republica, Facultad de Humanidades y Ciencias, Cerrito 73, Montevideo, Uruguay, South America
Arthur J. Riopelle, Department of Psychology, Emory University, Georgia
*Irwin L. Rodman, 2450 W. 18th Street, Los Angeles 6, California
Merrill Roff, University of Minnesota, Institute of Child Welfare, Minneapolis 14, Minnesota
Robert C. Rogers, R.F.D. No. 1, Princeton, New Jersey
John Rohrs, 15 Tamarack Road, Staten Island 4, New York
*Catherine M. Rosen, 1425 University Terrace, Ann Arbor, Michigan
Joseph R. Royce, Department of Psychology, University of Redlands, Redlands, California
Phillip J. Rulon, Peabody House, 13 Kirkland Street, Cambridge 38, Massachusetts
J. Francis Rummel, School of Education, University of Oregon, Eugene, Oregon
E. A. Rundquist, 1405 Summit Avenue, Alexandria, Virginia
*Jean M. Rutherford, 1205 Harrison Street, Syracuse, New York
T. A. Ryan, Department of Psychology, Cornell University, Ithaca, New York
David G. Ryans, Department of Education, University of California, Los Angeles 24, California
Richard T. Sampson, P. O. Box 296, Delaware, Ohio
Frank Sandon, 142 Rotton Park Road, Edgbaston, Birmingham 16, England
George A. Satter, Psychology Department, University of Michigan, Ann Arbor, Michigan
David R. Saunders, Cold Soil Road, R.F.D. No. 2, Princeton, New Jersey
Douglas E. Scates, Division of Teacher Education, 500 Park Avenue, New York 22, New York
Willis C. Schaefer, Institute for Research in Human Relations, 1156 19th Street, N. W., Washington 6, D. C.
*Lowell Schipper, Department of Psychology, University of Wisconsin, Madison 6, Wisconsin.
John Schmid, Personnel Research Laboratory, Lackland Air Force Base, San Antonio, Texas
*Kermit K. Schooler, Survey Research Center, University of Michigan, Ann Arbor, Michigan
William B. Schrader, Educational Testing Service, Princeton, New Jersey

Douglas Schultz, Educational Testing Service, Princeton, New Jersey
*Robert Schweiker, 1039 Vine Avenue, Williamsport, Pennsylvania
Saul B. Sells, 306 Abiso Avenue, San Antonio 9, Texas
Robert B. Selover, Prudential Insurance Company of America, Newark 1, New Jersey
*Hanan C. Selvin, 15 Kinkaid Avenue, Closter, New Jersey
Dr. Richard B. Seymour, 2026 Lake Street, Lincoln 2, Nebraska
Marion F. Shaycoft, 406 N. Neville Street, Pittsburgh 13, Pennsylvania
Joseph Russell Sherlock, 2685 Briggs Avenue, New York 58, New York
R. H. Shevenell, Institute of Psychology, University of Ottawa, Ottawa, Canada
Walter A. Shewhart, Mountain Lakes, New Jersey
Nathan Shock, Baltimore City Hospitals, Unit on Gerentology, Baltimore 24, Maryland
Herbert S. Sichel, Operational Research Bureau, P. O. Box 8152, Johannesburg, South Africa
Paul S. Siegel, Department of Psychology, University of Alabama, University, Alabama
George B. Simon, 808 S. Grove Street, Urbana, Illinois
Anthony J. Smith, Department of Psychology, University of Kansas, Lawrence, Kansas
J. E. Keith Smith, 719 Henry Street, Ann Arbor, Michigan
Max Smith, 453 E. 96th Street, Brooklyn 12, New York
John W. Stafford, Department of Psychology & Psychiatry, Catholic University of America, Washington, D. C.
John M. Stalnaker, 1075 Elm Street, Winnetka, Illinois
Julian C. Stanley, Jr., Department of Psychology, George Peabody College, Nashville 4, Tennessee
Artie John Stauffer, 754 Catalina Boulevard, San Diego, California
Frederick F. Stephan, Box 337, Princeton, New Jersey
Jack J. Sternberg, Walker House, 40 Quincy Street, Cambridge 38, Massachusetts
Naomi Stewart, 13 Doering Way, Cranford, New Jersey
Edward K. Strong, Jr., 672 Mirada Avenue, Stanford, California
Mr. Richard H. Sullivan, Box 592, Princeton, New Jersey
*Marcella Ann Sutton, 4816 Linscott Place, Apt. No. 3, Los Angeles 16, California
Frances Swineford, Educational Testing Service, Princeton, New Jersey
*Nancy A. Symons, Personnel Research Institute, 314 Superior Avenue, Cleveland 14, Ohio
Arthur Thomas Tait, 3746 Lankershim Boulevard, Los Angeles 28, California
Alfonso Tallet-Grenier, Universidad de Oriente, Santiago de Cuba, Cuba
John P. Tammaro, 281 Denison Parkway, Corning, New York
*Maurice M. Tatsuoka, c/o Mrs. Dorothy M. Glaser, 37 Hawthorne Street, Cambridge, Massachusetts
Calvin W. Taylor, Psychology Department, University of Utah, Salt Lake City, Utah
Erwin K. Taylor, 5530 Warwick Place, Chevy Chase 15, Maryland
Harold C. Taylor, W. E. Upjohn Institute For Community Research, 709 S. Westnedge Avenue, Kalamazoo 4, Michigan
Mahlon V. Taylor, 123 Spring Grove Road, Pittsburgh 35, Pennsylvania
Howard Tempero, Nebraska State Teachers College, Wayne, Nebraska
Lewis M. Terman, 761 Dolores Street, Stanford University, California
James A. Thomas, 625 E. Second Street, Xenia, Ohio
Godfrey Thomson, Moray House, University of Edinburgh, Edinburgh, Scotland
Kenneth F. Thomson, 25 Burbank Street, Yonkers 2, New York
Robert L. Thorndike, Teachers College, Columbia University, New York 27, New York
Robert M. Thrall, Department of Mathematics, University of Michigan, Ann Arbor, Michigan
L. L. Thurstone, Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina

Thelma G. Thurstone, Psychometric Laboratory, University of North Carolina, Chape Hill, North Carolina
David V. Tiedeman, Peabody House, 40 Quincy Street, Cambridge 38, Massachusetts
Kalman Tillem, 19 D Ridge Road, Greenbelt, Maryland
Helen Tomlinson, 1614 W. Rosewood, San Antonio, Texas
Herbert Toops, 1430 Cambridge Boulevard, Columbus 8, Ohio
Warren S. Torgerson, Educational Testing Service, Princeton, New Jersey
*Donald V. Torr, 2161 Waverly, San Antonio, Texas
Arthur E. Traxler, 21 Audubon Avenue, New York 32, New York
James B. Trump, 2814 74th Avenue, Apt. 202, Hyattsville, Maryland
Anthony C. Tucker, 211 Gonzalez Drive, San Francisco 27, California
Ledyard R. Tucker, 102 Cedar Lane, Princeton, New Jersey
William Watson Turnbull, 15 University Place, Princeton, New Jersey
Dik Warren Twedt, Young & Rubicam, Incorporated, 333 N. Michigan Avenue, Chicago, Illinois
Ralph W. Tyler, Department of Education, University of Chicago, Chicago 37, Illinois
C. C. Upshall, 78 Oneta Road, Rochester 5, New York
*Steven G. Vandenberg, 921 West Cross, Ypsilanti, Michigan
Neil J. Van Steenberg, 5556 N. 16th Street, Arlington, Virginia
Marvin J. Van Wagenen, University of Minnesota, Minneapolis 14, Minnesota
*Raymond H. Van Zelst, 206 Woodside Drive, Riverside, Illinois
C. L. Vaughn, Psychology Corporation, 522 5th Avenue, New York, New York
*Sylvia Vopni, 2107 28th Avenue West, Seattle 99, Washington
David F. Votaw, Box 587, San Marcos, Texas
Helen M. Walker, Teachers College, Columbia University, New York 27, New York
William B. Walker, 804 Louise Circle, Durham, North Carolina
Edwin Wandt, 1832 Westholme Avenue, Los Angeles 25, California
M. J. Wantman, Catherine Strong Hall, University of Rochester, Rochester 3, New York
Neil D. Warren, 3827 Olympiad Drive, Los Angeles 43, California
R. R. G. Watt, 1188 W. 28th Street, Los Angeles, California
Herbert B. Weaver, Department of Psychology, University of Honolulu, Honolulu, Hawaii
David Wechsler, 145 E. 92nd Street, New York 28, New York
Alexander G. Wesman, 84 Ralph Avenue, White Plains, New York
Benjamin N. White, 715 Lawrence Street, Ann Arbor, Michigan
*Dean K. Whitla, 1747 P Street, N. W., Washington, D. C.
Alfred G. Whitney, Life Insurance Agency Management Association, 855 Asylum Avenue, Hartford, Connecticut
Lee M. Wiggins, 135 W. 56th Street, New York 19, New York
S. S. Wilks, Princeton University, Princeton, New Jersey
Clark L. Wilson, Jr., 721 W. Jefferson Boulevard, Los Angeles, California
Ben James Winer, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina
John F. Winne, Department of Psychology, V. A. Hospital, Northampton, Massachusetts
J. R. Wittenborn, 85 Edgemere Road, Hamden, Connecticut
Herman O. A. Wold, Institute of Statistics, Odinsland 2, Uppsala, Sweden
Dael L. Wolfle, 2801 Rittenhouse Street, N. W., Washington 15, D. C.
Ben D. Wood, Columbia University, New York 27, New York
David L. Wright, R.D. No. 4, c/o Taylor, Greensburg, Pennsylvania
J. W. Wrightstone, 21 Hickory Road, Summit, New Jersey
C. F. Wrigley, Department of Psychology, University of Illinois, Urbana, Illinois
Virginia Zachert, 203 McMullin Street, San Antonio 10, Texas
Robert C. Ziller, 223 E. Jackson, Colorado Springs, Colorado
J. Zubin, 722 W. 168th Street, New York, New York

INDEX FOR VOLUME 17

AUTHOR

Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY · ORGANIZED IN 1935

VOLUME 17
NUMBER 4
DECEMBER
1952

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

DECEMBER 1952 VOLUME 17, NUMBER 4

Printed for the Psychometric Society at 1407 Sherwood Avenue, Richmond 5, Virginia. Entered as second class matter at the Post Office of Richmond, Virginia. Editorial Office, Department of Psychology, The University of North Carolina, Chapel Hill, North Carolina.

Subscription Price: The regular subscription rate is $10.00 per volume. The subscriber receives each issue as it comes out, and a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. All back issues are available. Back issues are $10.00 per volume (one set only) or $2.50 per issue, with a 20 per cent discount to Psychometric Society members. Members of the Psychometric Society pay annual dues of $7.00, of which $6.30 is in payment of a subscription to *Psychometrika*. Student members of the Psychometric Society pay annual dues of $4.00, of which $3.60 is in payment for the journal.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

HARRY H. HARMAN
4111 Maryland Drive, Washington 16, D.C.

Payments: All bills and orders are payable in advance. Checks covering membership dues should be made payable to the *Psychometric Society*. Checks covering regular subscription to *Psychometrika* and back issue orders should be made payable to the *Psychometric Corporation*. All checks, notices of change of address, and business communications should be addressed to

ERWIN K. TAYLOR, Treasurer, Psychometric Society and Acting Treasurer, Psychometric Corporation
Personnel Research Institute, Western Reserve University
Cleveland 14, Ohio

Articles on the following subjects are published in *Psychometrika:*

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), in so far as articles of this type are available.

In the selection of the articles to be printed in *Psychometrika*, an effort is made to obtain objectivity of choice. All manuscripts are received by one person, who first removes from

(*Continued on the back inside cover page*)